

PB85-169191

National Urban Pesticide Applicator Survey
Final Report. Technical Volume 2
Data Processing, Editing, and Weighting

Research Triangle Inst.
Research Triangle Park, NC

Prepared for

Environmental Protection Agency, Washington, DC

13 Apr 84

U.S. Department of Commerce
National Technical Information Service

NTIS

April 13, 1984

RTI/2764/08-01F

NATIONAL URBAN PESTICIDE APPLICATOR SURVEY:
FINAL REPORT

Technical Volume II: Data Processing, Editing, and Weighting

by

Frederick W. Immerman
Lanny L. Piper
Douglas J. Drummond
Jeanne R. Pagano
Ralph E. Folsom
Sara C. Wheelless

Research Triangle Institute
Research Triangle Park, NC 27709

Contract No. 68-01-6646

Task Manager: Linda Zarow
Project Officer: Edward Brandt

Economic Analysis Branch
Benefits and Use Division
Office of Pesticide Programs
U.S. Environmental Protection Agency
401 M. Street, SW
Washington, DC 20460

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

REPORT DOCUMENTATION PAGE	1. REPORT NO.	2.	3. Recipient's Accession No. PB8 5 169191 /AS
4. Title and Subtitle National Urban Pesticide Applicator Survey: Final Report Technical Volume II; Data Processing, Editing & Weighting			5. Report Date April, 1984
7. Author(s)			6.
9. Performing Organization Name and Address Research Triangle Institute Post Office Box 12194 Research Triangle Park, North Carolina 27709			8. Performing Organization Rept. No. 54019-84-008
12. Sponsoring Organization Name and Address U.S. Environmental Protection Agency Office of Pesticide Programs Benefits & Use Division (TS-768-C) 401 M Street SW, Washington, D.C. 20460			10. Project/Task/Work Unit No. Same as #11
			11. Contract(C) or Grant(G) No. (C) RTI/2764/08-01F (G)
15. Supplementary Notes This report was prepared for the U.S. Environmental Protection Agency under contract to Research Triangle Institute and sponsored by the four trade associations listed in item #16. This is a final report.			13. Type of Report & Period Covered Final Report
16. Abstract (Limit: 200 words) A study has just been completed of pesticide usage in three sectors of the commercial pest control industry-tree, lawn and structural. This study was funded by the EPA and was jointly sponsored by the three trade associations representing these sectors and the Association of American Pesticide Control Officials. A total of 2,838 firms in the industry responded to the survey by providing data for the year 1981. They represented all regions of the country, and all firm sizes. The sample included firms belonging to the cooperating trade associations and other firms as well. About one out of every three eligible businesses from individual state licensing lists was included in the sample. All states except North Dakota and Alaska participated. Sample selection was in accordance with a probability-based design permitting valid statistical inferences for each industry sector and geographical region. Respondent eligibility criteria are indicative of the scope of the survey. A business was eligible if it--1) was on a state licensing list, 2) was in business in 1981, 3) provided commercial pest control services, and 4) did not provide strictly agricultural pest control services.			14.
17. Document Analysis a. Descriptors			
b. Identifiers/Open-Ended Terms			
c. COSATI Field/Group			
18. Availability Statement "Release unlimited"	19. Security Class (This Report) Unclassified	21. No. of Pages 154	
	20. Security Class (This Page) Unclassified	22. Price	

TABLE OF CONTENTS
Technical Volume II

	<u>Page</u>
ACKNOWLEDGEMENTS.	
1. INTRODUCTION	1-1
1.1 Overview of Technical Volume II	1-1
1.2 Schedule of Activities.	1-2
2. INITIAL DATA PROCESSING	2-1
2.1 Hard Copy Edit	2-1
2.2 Keying of Questionnaire Data	2-2
3. TREATMENT OF INDUSTRY "GIANTS".	3-1
3.1 Rationale for Special Treatment.	3-1
3.2 Response from the National Offices	3-3
3.3 Identification of Frame and Sample Representatives of the Industry Giants.	3-4
3.4 State-Level Estimates.	3-5
4. WEIGHTING	4-1
4.1 Overview	4-1
4.2 Firm-Level Basic Weights	4-2
4.2.1 Overview.	4-2
4.2.2 Mail States	4-3
4.2.3 Telephone States.	4-3
4.3 Multiplicity	4-5
4.3.1 Sources of Multiplicity	4-5
4.3.2 Identifying Multiplicities.	4-6
4.4 Adjustment for Industry Giants	4-8
4.5 Nonresponse Subsampling.	4-9
4.6 Determining Eligibility.	4-12
4.6.1 Using Questionnaire Data.	4-12
4.6.2 Undeliverable/Unable to Contact Adjustment.	4-12
4.7 Defining Respondents	4-15
4.8 Defining Nonresponse Weighting Classes	4-15
4.9 Nonresponse Adjustments.	4-19
4.9.1 Multiplicity Adjustment for Nonrespondents.	4-19
4.9.2 Adjustment for Refusals	4-21
4.9.3 Weighting Class Nonresponse Adjustment.	4-23
5. VARIANCE ESTIMATION	5-1
5.1 Variance Estimation in Complex Surveys	5-1
5.2 Variance Estimation Formula.	5-3
5.3 Variance Estimation Software	5-7
5.4 Construction of the Usage Analysis File.	5-9

TABLE OF CONTENTS (continued)

Technical Volume II

	<u>Page</u>
6. USAGE DATA CLEANING AND EDITING	6-1
6.1 Intended Utilization of Product Usage Data	6-1
6.2 Characteristics of the Raw Usage Data.	6-2
6.3 Assignment and Editing of Registration Numbers	6-4
6.4 Resolution of Nonspecific Unit Codes	6-7
6.5 Amount Imputation.	6-13
6.6 Product Class and Formulation Categorization	6-14
6.7 Estimation of Active Ingredient Usage.	6-15
7. EDITING NONUSAGE DATA FOR LOGICAL CONSISTENCY	7-1
7.1 Intended Level of Editing.	7-1
7.2 Editing of Specific Questions.	7-2
8. EXTERNAL REVIEW OF PRELIMINARY RESULTS.	8-1
8.1 Motivation for External Review	8-1
8.2 Results of the Review Process.	8-1
9. DATA QUALITY.	9-1
9.1 Realized Target Population	9-1
9.2 Coverage Rates	9-3
9.3 Potential Sources of Error	9-3
9.4 Quality of the Usage Data.	9-8
9.5 Quality of the Nonusage Data	9-12
9.6 Overall Strengths and Weaknesses of the NUPAS Data	9-13
10. THE ANALYSIS PLAN FOR NUPAS DATA.	10-1
APPENDIX A Edit Manual for NUPAS	A-1
APPENDIX B Reporting Aggregate Resolution Questionnaire.	B-1
APPENDIX C: Letter to Dr. Richard Carr Outlining Information Required for Interpretation of Certification/Licens- ing Questions and Definition of Realized Target Population.	C-1
APPENDIX D Memo to EPA on the NUPAS Analysis Plan and User File Content	D-1

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1-1	Originally Anticipated Schedule of Volume II Activities. .	1-2
1-2	Realized Schedule of Volume II Activities.	1-3
3-1	Event Status Codes for Sampled Members of Industry Giants, as of July 31, 1982.	3-2
6-1	Initial Meanings for Edit Flag Values.	6-6
6-2	Revised Meanings for Edit Flag Values.	6-8
6-3	Unit Codes for Nonspecific Packaging Information	6-11
6-4	Definition of NUPAS Product Class Categories	6-16
6-5	Definition of NUPAS Product Formulation Categories	6-17
8-1	Reviews and Presentations of Preliminary NUPAS Results . .	8-2
9-1	Coverage Rates for Eligible Respondents Prior to the Nonresponse Followup Subsampling	9-4
9-2	Coverage Rates for Eligible Respondents After the Non- response Followup Subsampling.	9-5
9-3	Coverage Rates for Eligible Respondents Who Provides Usage Data, After the Nonresponse Followup Subsampling	9-6
9-4	Amount of National Usage of 10 Active Ingredients at Four Steps in the Usage Estimation Procedure.	9-11

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
5-1	Flow of Logic for the Variance Estimation Software	5-8
6-1	Usage Data Status.	6-9
6-2	Usage Data Status.	6-10

ACKNOWLEDGEMENTS

RTI wishes to acknowledge the efforts of the following persons who were involved with the NUPAS activities described in this volume.

Association of American Pesticide Control Officials

Barry Patterson, Director,
Division of Agricultural and Environmental Services,
N. M. Department of Agriculture
Arlo Ehart, Chief, Pesticide Use and Control, Plant Industry
Division, Wisconsin Dept. of Agriculture

Association of Structural Pest Control Regulatory Officials

Neil Ogg, President
All members in attendance at the 1983 annual meeting

Chemlawn Corporation

Robert W. Miller, Vice President, Support Services

Consultants

Richard Carr, Urban Biosystems, Inc.
Edwin Schneider

The Davey Tree Expert Company

Donald J. Shope, Vice President and General Manager,
Tree Care Services

Environmental Protection Agency

Arnold Aspelin, Chief, Economic Analysis Branch
Edward Brandt, Economist, Economic Analysis Branch
Peter Kuch, Supervisory Economist, Economic Analysis Branch
Linda Zarow, Economist, Economic Analysis Branch
David Brussard, Entomologist, Animal Biology Information Section

The F.A. Bartlett Tree Expert Company

Walter E. Dages, Public Relations Director

National Arborist Association

Robert Felix, Executive Vice President

National Pest Control Association

Jack Grimes, Director, Government Affairs
Members of the Project Development Council in attendance
at the 1983 annual meeting

North Carolina Agricultural Extension Service

James Baker, Entomologist

Orkin Pest Control

Robert M. Russell, Vice President, Government Relations

Research Triangle Institute

Douglas Drummond, Project Director

Center for Survey Research

Vivian Adkins
William Chapman
Charlene Derossett
Sandra Fox
Linda Miller
Evelyn Scalf
Richard Waddell

Center for Survey Statistics

Francine Burt
Martha Clegg
Tim Gabel
Vince Iannacchione
Laurine Johnson
Phyllis Norris
Lisa Packer
Pat Parker
Brenda Porter
Mary Ann Rowland
Scott Sweetland

Center for Computer Applications

Danny Allen
Thomas Farrell
Brenda Frick
Beverly Hitchcock
David Moazed
S. Kay Powell
Mary Beebe Woodside

Research Services Department

Robert Nichols
Martha Roberts
Elizabeth Weaver
Duplicating & Reproduction Staff
Telephone Survey Unit Staff

Terminix International, Inc.

John A. Craft, Technical Director

Union Carbide Agricultural Products

David Herman, Market Analyst

Velsicol Chemical Corporation

Charles Frommer, Director, Regulatory Affairs

Whitmire Research Laboratories

Daniel Stout

TECHNICAL VOLUME II

1. INTRODUCTION

1.1 Overview of Technical Volume II

This volume of the NUPAS final report describes how the NUPAS data was treated prior to the analyses and tabulations which are presented in the Overview and Results Volume. The intent is to provide the information necessary for a knowledgeable, accurate interpretation of the numerical results. The topics covered in this volume include:

- cleaning and editing of the pesticide usage data
- editing of the nonusage data
- generation of the analysis weights
- the variance estimation procedures used for NUPAS data
- assessments of data quality.

Technical Volume I of the NUPAS report provides the foundation for much of the discussion in this volume, and will be referred to repeatedly.

1.2 Schedule of Activities

The activities described in this volume occurred primarily in the period from October 15, 1982 through February 29, 1984. The original schedule for these activities is shown in Table 1-1. By November, 1982, RTI realized that the data quality and reporting characteristics were such that adherence to the original schedule would be impossible if accurate estimates were to be produced. In particular, cleaning of the usage data, resolution of the reporting aggregates, and calculation of analysis weights required considerable adjustment of schedule. The realized schedule for Volume II activities is given in Table 1-2.

Table 1-1. Originally Anticipated Schedule of Volume II Activities

<u>Activity</u>	<u>Completion Date</u>
Edit questionnaires	October 15, 1982
Place questionnaires in computer accessible form	October 25, 1982
Construct analysis weights	November 30, 1982
Convert reported usage data into pounds of active ingredients	December 30, 1982

Table 1-2. Realized Schedule of Volume II Activities

Activity	Completion Date
Initial editing of questionnaires	October 15, 1982
Place questionnaires in computer-accessible form	November 1, 1982
Obtain national data from industry "giants"	March 1, 1983
Resolve composition of reporting aggregates	June 1, 1983
Convert reported usage data into pounds of active ingredient	December 1, 1983
Logical editing and imputation of non-usage data	January 15, 1984
Construct analysis weights	January 30, 1984
Develop variance estimation methodology	January 30, 1984

2. INITIAL DATA PROCESSING

2.1 Hard Copy Edit

RTI project staff met in early April 1982 to begin preliminary planning for the manual edit and data entry processes. Before work could begin on the data entry program, acceptable ranges for coded responses had to be established. Likewise, certain manual edit decisions had to be formulated before any data entry planning could begin. These involved general edit guidelines and rules, specific "key" question handling procedures, what to do if a questionnaire were to "fail edit"--in other words, one or more "key" questions not answered, and handling of other minor problems that could occur. Once a draft of the edit specifications was completed, it was reviewed by RTI staff as well as the EPA Project Officer. Upon approval, it was reproduced in preparation for training of the edit staff. A copy of the edit manual is reproduced in Appendix A.

The project survey specialist and survey assistant conducted a thorough training session on the edit procedures; editors were selected from RTI's Research Services Department. They were supervised by one of their group supervisors who was also in attendance at the training (she was briefed on the edit and supervisory procedures before the general training). The training included actual edits of several completed questionnaires; any problems encountered during this practice session were discussed by the trainers and the entire group. Following the training, the survey staff remained with the editors for several days to answer questions, resolve problems, etc. Once they felt sure that everyone understood the procedures, the data services supervisor took charge.

Any problems with questions that needed resolution were documented on the Edit Problem Sheet and referred to the supervisor. She would either resolve the problem(s) herself or refer them to project staff, as necessary. Once questionnaires had completed the edit process, they were re-batched and held for data entry.

2.2 Keying of Questionnaire Data

Following the hard copy edit, the questionnaires were keyed onto a computer file by data entry specialists. Quality control was enhanced by the use of data entry software that performed range checks for each variable keyed. All questions asked in the NUPAS questionnaire were represented on the computer file. Response to the questions on problems encountered and contact person were keyed only as having information present or absent. Respondent's notes in the "other" sections of the questions on trade association membership, magazine subscriptions, sites served, and industry sector were not keyed due to schedule and format considerations. All usage data was represented in exactly the form to which it had been edited during the hard copy edit process.

A sample of records in each batch were checked for keying accuracy by checking against double re-keys (i.e. comparison of three copies). The average character error rate by batch was approximately 0.3%, and no batch exceeded the RTI maximum acceptable limit of 0.5%. The product name and EPA registration number fields were not included in the calculation of batch error rates because all records in both of these fields would be examined and updated during the process of product data cleaning.

Because the usage data required extensive cleaning and editing, for ease of processing, the questionnaire file was divided into two files following verification. One file contained a single record for each questionnaire comprised of all nonusage variables. The other file, hereafter

referred to as the usage file, contained each of the variables from the product usage question (question 22, see Appendix A of the Overview and Results Volume), as well as the questionnaire identification number. Each line of question 22 was represented as a single record on this file.

3. TREATMENT OF INDUSTRY "GIANTS"

3.1 Rationale for Special Treatment

In any survey in which a very small proportion of the target population controls a large proportion of the quantity to be measured, the accuracy of the estimates produced can be adversely effected if the data for that small segment is not of good quality. During the planning stages of NUPAS it was recognized that three companies -- Orkin Pest Control, Terminix, and Chemlawn, henceforth referred to as the industry "giants" --accounted for a substantial proportion of the pesticide usage in their respective sectors, and that it would be important to obtain high quality data from them. Each of these firms provides service through a large number of locations which together cover much of the continental United States.

Originally it was felt that special treatment of these firms in the NUPAS sample design was unnecessary and undesirable. By treating industry giant locations in the same way that other frame members were handled, analysis of the data would be simplified and the giants would be assured of confidentiality. Thus, multiple locations were sampled for each of the giants, and each received the standard NUPAS questionnaire.

By mid-July 1982, it was apparent that the response from the sampled locations of the giants was not adequate to provide a sufficiently high quality estimate of their pesticide usage (see Table 3-1). The sampled locations that did respond often indicated that portions of the questionnaire were difficult or impossible for them to fill out; in some cases, they cited a lack of records or the necessity of national headquarter's permission to release the requested data.

Table 3-1. Event Status Codes for Sampled Members
of Industry Giants, as of July 31, 1982

Event Status Code (x=1 or 2)	Description	Number of Sample Members		
		Wave 1	Wave 2	Overall
90	Sampled, not mailed	8	-	8
x00	In the field	0	183	183
x02	Undeliverable	13 ¹	0	13
x06	Ineligible	3	0	3
x08	Partial respondent	6	17	23
x09	Partial respondent (branch)	0	1	1
x10	Complete respondent	<u>4</u>	<u>4</u>	<u>8</u>
		34	205	239

¹Some of these may have been due to the fact that home, rather than business, addresses were reported for some elements on some state lists.

Because high quality data for the giants was so important, the survey team decided to contact the national offices of each and ask for aggregated, national-level responses to the questionnaire. For weighting and analysis purposes, the national data for these giants was treated as census data, rather than a sample. Further discussion of the impact and implications of this special treatment occurs in section 4.4 and chapter 9.

Several other large companies were asked for national, aggregated data following the reporting aggregate resolution callback effort (see section 4.3). Although these companies are large, they are either not the dominant companies in their sector, or do not use large quantities of pesticide compared to the three industry giants discussed above. Therefore, the data from these firms are not treated as census data, but rather as sampled reporting aggregates with multiplicity.

3.2 Response from the National Offices

Following several contacts in which the importance of their participation was stressed and assurances of confidentiality made, the national headquarters of each of the three industry giants responded to the request for national data. The national NUPAS usage estimates are undoubtedly more accurate than they would have been without this cooperation.

Orkin and Chemlawn provided completed versions of the questionnaire, answering all applicable questions with aggregate, national data. Chemlawn distinguished in its reporting of product usage between the pesticides used by their Chemlawn division and those used by their Chemscape division. After further contact, the information needed to apportion the answers to the nonusage questions between the two divisions was also provided. In all analyses, Chemlawn and Chemscape are handled as two distinct industry giants.

Terminix provided national answers to some of the nonusage questions and provided computer printouts for each of its branch offices, indicating the amounts of pesticides purchased by each from Terminix. Each branch office serves as a local distribution center for from 4 to 30 Terminix locations. On the printouts a distinction was made between company-owned offices, which are required to purchase all of their chemicals from Terminix, and franchises, which are free to purchase pesticides from other vendors as well. For the NUPAS survey, Terminix franchises were included in the sample portion of NUPAS, and the company-owned offices in the census portion. Therefore, usage data for all company-owned offices are included in the analyses, while data for the franchises were used only for those locations that were sampled. Data from franchises that responded on their own to the NUPAS questionnaire were retained, and, if necessary, the usage data for these samples were supplemented by data from the national headquarters. Usage estimates for Terminix franchises that did not respond on their own may be low since these franchises could have purchased additional products from distributors other than Terminix. However, on the information available to them, Terminix headquarters estimated that the degree of such underreporting was minor.

3.3 Identification of Frame and Sample Representatives of the Industry Giants

To avoid double counting of industry giant locations that responded to the original sampling effort, and to calculate more efficient analysis weights, all of the sample members and frame members associated with the industry giants had to be identified as such. In the telephone states, identification of all industry giant frame members was not possible due to the lack of firm identification information on the first-stage frame.

Identification of Orkin, Chemlawn, and Chemscape members of the sample and the mail state frame was relatively straightforward. However, for Terminix, a number of factors made the mapping process more difficult. Terminix locations were identified by a variety of names on the frame (e.g., Bruce-Terminix, Terminix International, Terminix Services), making computerized searching for name matches less efficient. Sears Termite and Pest Control is also often (but not always) a Terminix operation, and these were covered by the usage data provided by Terminix. Terminix franchises had to be distinguished from Terminix company-owned offices, although in many cases the address on the frame did not match any of the branch office addresses provided by Terminix. This was due in part to inaccurate or unstandardized address listings on the state lists, and in part to the fact that some states regulated at levels higher or lower than the branch office level. This was particularly true in California, which regulates firms only at a relatively high level (i.e., district office or company headquarters). All of these difficulties were ultimately overcome through repeated contacts with the Terminix head office. Overall, 626 frame units in the mail states and the mail portion of Louisiana were associated with the giants, and 207 were sampled; in the telephone states, 35 phase II sampled applicators worked for the industry giants.

3.4 State-Level Estimates

To use national data provided by the industry giants for analyses at the state or regional level, it was necessary to devise methods of prorating the responses to some questions across the states served by the company. For Orkin, Chemlawn, and Chemscape data, the number of employees, number of accounts, and pesticide usage amounts were prorated based on the percent of the company's overall sales in the state or region of interest.

This method was chosen because the giants were willing and able to supply the necessary information, and because it was relatively easy to incorporate in the analyses. To the degree that the distributions of employees, accounts, and pesticide use depart from the relative distribution of gross sales, this method of proration creates a source of error in the estimates. This may be especially true for usage estimates because some products are used only in certain geographic areas.

Because Terminix provided usage data at the branch office level, there was only limited need to prorate those data. For Terminix offices serving more than one state, usage was prorated based on the relative breakdown of the office's business in the states it served. Terminix provided national data on the number of employees in each of the categories listed in question 10. Because each Terminix company office was represented separately on the analysis file, the total (national) number of employees had to be prorated across the offices. The only readily available measure for this proration was the amount of pesticide (pounds) used by an office relative to the total amount of pesticides used by all offices. This method was used despite its potential for error, since the alternative was to not use the employee data at all.

It should be noted at this point that none of the giants could provide a usable response to the questions on the number of applicators licensed or certified in the reference state. This was due in part to variation among states in the definitions of the terms "licensed" and "certified," and in part due to a lack of such information at the national offices. Therefore, any analyses of the responses to these questions do not include any data from the industry giants.

4. WEIGHTING

4.1 Overview

Sample weighting is a visible recognition by the research scientist that the data provided by respondents should not always have a uniform influence in the analysis. Many factors can affect the need for sampling weighting including inefficiencies of the sampling frame, unequal probabilities of selection applied to the elements of the sampling frame, stratification of the sampling frame, level at which the respondent chooses to report, and nonresponse. The sample weighting procedures for NUPAS recognized these survey features and embodied them in the sampling weights.

In its simplest form a sampling weight can be described as the number of units on the sampling frame represented by the associated sampled unit. A basic sampling weight in a probability sample is computed as the inverse of the expected number of times a unit gets selected into the sample. Adjustment of the basic weight can then be made in an attempt to account for events and circumstances that disrupt the original sample design and associated inclusion probabilities.

Three major features of NUPAS made it necessary to adjust the basic weights. The first was that some respondents provided data for more than one frame unit. This is termed multiplicity, and was handled by using multiplicity estimators. The second factor, already discussed in Chapter 3, was the special treatment afforded three industry giants. Finally, adjustment had to be made for the nonresponse which remained following the personal interview followup effort. Because there were no externally-derived standards to adjust to, nonresponse adjustments were based on weighting class sampling frame totals.

4.2 Firm-Level Basic Weights

4.2.1 Overview

The basic weights for this survey were constructed to reflect the probability mechanism that was used to select the initial sample of firms. The sampling frame was compiled with firms as the sampling unit and the data requested by the survey instrument was firm-level information with some geographic and firm name restrictions. For these reasons firm-level weights would be the most direct and the least complicated set of weights that could be computed.

Certain underlying assumptions inherent in the formulation of the basic weights must be recognized. Deviations from these assumptions make it necessary to adjust the basic weights. Deviations in NUPAS were caused by inefficiencies in the sampling frame, inability or unwillingness of the respondent to respond at the level requested, and total nonresponse by some sample members. The basic weights assume that each firm is listed only one time on the frame and thus can be included in the sample only by selecting the one listing. When this is not the case, multiplicity is introduced. Multiplicity occurs when more than one listing on the frame is linked to the ultimate respondent, and was anticipated from the beginning in this survey due to the nature of the sampling frame.

Multiplicity can also be introduced by the respondent when the data reported does not correspond to the level of the sampling unit. For this survey, the data was requested for the named firm's business in a specified state. If that firm chose to report its business in all the states it serves, and if the firm was listed on the sampling frame in at least one of the other states served, then the respondent has introduced a multiplicity.

As discussed in Chapter 2, Technical Volume I, two distinct sample designs were required to achieve a firm-level sample. A single-stage list sample was applied in 42 of the states and a portion of Louisiana, while a two-phase list sample design was applied in six states and the remainder of Louisiana using telephone screening in the first phase of the design. Since the weights reflect the probability mechanism used in the selection process, the steps taken to construct the basic weights for the mail states differ from the steps taken to construct them for the telephone states. The differences will be described in the following sections.

4.2.2 Mail States

The mail state sample was composed of independent samples of firms within each state. For each mail state, a list of pesticide businesses was compiled. This list was then stratified by industry sector category (see Chapter 2, Technical Volume I). Within each state the firms listed in a stratum were assigned an equal probability of selection. The basic weight for each sampled firm was computed as the reciprocal of its selection probability. Therefore, all firms assigned to the same stratum in a given state had the same basic weight. Basic weights for firms on the 'mail' frame in Louisiana were computed in the same way.

4.2.3 Telephone States

For each telephone state (including part of Louisiana) a two-phase sample design was implemented. In the first phase a sample of pesticide applicators that were licensed by the state was selected independently from each of two strata. Therefore, in a telephone state, each sampled applicator in a given stratum had the same probability of selection. This sample of applicators was then telephoned (see Appendix A of Technical Volume I) to identify all NUPAS eligible firms that employed the sampled applicators.

The applicators who responded and identified eligible firms composed the second-phase sampling frame. These applicators were further stratified according to their firm's industry sector category. Applicators who identified more than one eligible firm were assigned to the stratum associated with their first listed firm. Within each telephone state, equal selection probabilities were assigned by stratum to the applicators in the second-phase frame. When an applicator was selected for the second-phase sample, all eligible firms listed by the applicator were included in the sample and thus the applicator's selection probability was associated with each firm he/she had listed. The resulting sample of firms was then mailed the same questionnaire that was mailed to the sample of firms in the mail states. Obviously, this process introduces multiplicity in the frame listing of firms whenever different applicators work for the same firm. This issue is addressed in Section 4.3.

The basic weight for sampled firms in the telephone states consists of components from both the first and second-phase samples. The first-phase component is the reciprocal of the applicator's selection probability adjusted for applicator nonresponse and for eligible applicator nonresponse. The adjustment had to be handled in two steps because eligibility could not be directly determined for nonrespondents. The applicator nonresponse adjustment was computed within each state by stratum cell as the ratio of the number of applicators selected in the cell to the number of responding applicators in the cell. It was then determined whether or not those responding applicators worked for a NUPAS eligible firm. The nonresponse adjustment for eligible applicators was then computed for each state by stratum cell as the ratio of the number of applicators determined to be working for an eligible firm to the number of those applicators who provided the firm's name. The component of the basic weight representing the second-phase of sampling was the reciprocal

of the applicator's second-phase probability of selection. The basic weight for applicator's selected in the telephone states was then generated as the product of the first-phase component, the applicator nonresponse adjustment, the eligible applicator nonresponse adjustment, and the second-phase component. Algebraically, the basic weight for applicators in telephone states can be represented as:

$$T_{hijk}^W = \frac{1}{1\pi_{hi}} \cdot \frac{n_{hi}}{m_{hi}} \cdot \frac{a_{hi}}{b_{hi}} \cdot \frac{1}{2\pi_{hij}}$$

where T_{hijk}^W = the basic weight for applicator $hijk$ in the telephone states

$1\pi_{hi}$ = the first-phase selection probability for stratum- i in telephone state- h

$2\pi_{hij}$ = the second-phase selection probability for stratum- ij in telephone state- h

n_{hi} = sampled number of applicators in stratum- i of state- h for the first-phase sample

m_{hi} = the number of the n_{hi} applicators that responded

a_{hi} = the number of first-phase responding applicators reported to be working for a NUPAS eligible firm

b_{hi} = the number of the a_{hi} applicators that provided the eligible firm's name

4.3 Multiplicity

4.3.1 Sources of Multiplicity

As previously stated, multiplicity can be generated in the construction of the sampling frame by listing the same sampling unit more than once on the frame. Multiplicity can also be generated by a respondent who reports at a level that includes more than one sampling unit on the frame. Sometimes the units reported on the questionnaire include both units selected for the sample and units not selected for the sample.

The following situations were sources of multiplicity in the mail states:

- firms having multiple locations (addresses) within the same state
- firms with different names but having the same ownership and working out of the same location (address)
- firms who served and were licensed in more than one state
- firms that had some of their employees certified in one of the telephone states that they served.

Sources of multiplicity in the telephone states included:

- firms that employed more than one licensed pesticide applicator who were sampled and responded in phase one of the design
- firms that also were licensed in and served another state
- firms having employees licensed in more than one telephone state such as Virginia and West Virginia, or Pennsylvania and West Virginia.

4.3.2 Identifying Multiplicities

During the frame construction process RTI attempted to remove as many multiple listings of the same firm as possible. This process is described in Technical Volume I, Chapter 2. Part of this process included preparing lists of potential multiplicities within each state that included firms that were candidates to be removed from the frame but did not satisfy all the constraints for removal. These lists were used as the first check when a returned questionnaire indicated the potential for a multiplicity.

Questions 5, 6, 7, and 8 of the questionnaire (see Appendix A of the Overview and Results Volume) were used to set flags to indicate a potential for multiplicity. Respectively, these questions attempted to identify the number of business offices or locations of a firm in the reference state, the service area of the firm (multi-state or single state), the existence of firms with a different name but having the same ownership as the sampled firm, and whether the firm had undergone any splits or mergers with other firms in the

preceding 12 months. Another flag was used to indicate potential multiplicity when the reference state on a questionnaire was not the same as the state in the mailing address.

The questionnaire did not collect identification information from respondents that indicated their firm had multiple locations, or operated under more than one name or served multiple states or had split or merged with another pest control firm. This would have been no problem had the respondents not misinterpreted the instructions in the questionnaire that defined the level for which data was to be reported. The respondent was given the option of not responding if he/she could not provide data at the requested level. However, as the returned data was reviewed by analysts, it became obvious that the data reported was very often not solely for the requested firm's business in the given reference state. Because of this, it was necessary to recontact some respondents to clarify the relationship of the pesticide usage data reported to the firm-level information reported in questions 5, 6, 7, and 8.

A telephone callback operation was designed to confirm the level of data aggregation in questionnaires that indicated a potential for multiplicity. There were approximately 800 such questionnaires designated for this callback process. In addition to confirming the level of aggregation in the usage data, this callback also attempted to obtain the information needed to identify any frame elements and sample members included in the reporting aggregate (see Appendix B), and for those firms selected from a telephone state, or that served a telephone state or states, to verify the number of applicators certified in that telephone state.

The information obtained through the callback process was used to search the frame for multiplicities. When a multiplicity was confirmed on the frame

listing, the number of multiple frame listings was recorded on the firm's data record as an additional variable. A multiplicity value of one (1) was assigned to those firms in which potential multiplicities were not identified and firms with a potential multiplicity but for which the frame search for multiple listings proved unsuccessful.

The callback procedures also led to the identification, in some cases, and the discovery, in other cases, of firms that had returned more than one questionnaire. Some such firms had the same firm name while others had different names. In such cases, the data were not always consistent across the multiple instruments and the data from the callback was used in an attempt to resolve the inconsistency. Even at this level the inconsistency could sometimes not be resolved and thus an additional callback was necessary for final resolution. Additional callbacks were also required in certain other cases to resolve inconsistencies between the original data, the callback data and the frame data.

Those respondents who reported data for more than one state were asked to indicate the percent of their sales associated with each state reported on. These percents were used to prorate the usage and other data for state (and, in some cases, regional) analyses (see section 3.4).

4.4 Adjustment for Industry Giants

Chapter 3 of this document, a discussion was presented concerning the special handling of the industry giants, namely Terminix, Orkin, and Chemlawn/Chemscape. The procedures applied to the giants were implemented after the sampling frames were constructed and after the sample had been selected. The frames contained listings for Terminix, Orkin, and Chemlawn/Chemscape firms and each listing had its own selection probability. The decision to treat the original sample as a census of the industry giants and a sample of all other

firms prompted adjustment of the mail state basic weights to better estimate the frame size without the giants.

The first step in the adjustment was to identify frame listings associated with the giants. The total number of such listings was then determined for each stratum within each mail state. Similarly, the number of firms in the sample and associated with the giants was determined for each sampling stratum in each mail state. The basic weights for the nongiant firms were then computed follows:

$$M_{hi}^W = \frac{N_{hi} - M_{hi}}{e(n_{hi}) - m_{hi}}$$

where

M_{hi}^W = the basic weight of a firm selected from stratum-i of mail state-h

N_{hi} = the total number of firms listed on the sampling frame for state-h and stratum-i

M_{hi} = the total number of firms associated with the giants in state-h and stratum-i

$e(n_{hi})$ = the expected number of firms sampled from state-h and stratum-i out of N_{hi} ¹

m_{hi} = the number of firms sampled from state-h and stratum-i that were associated with the giants.

4.5 Nonresponse Subsampling

A sample of those firms that had not responded after a second mailing of the questionnaire was selected for personal followup by AARP. This process and its rationale are described Technical Volume I, chapters 2 and 6. This

¹The expected rather than the actual number is used because a probability minimum replacement algorithm was used to allocate the sample. See Chromy, J. R. (1979), "Sequential Sample Selection Methods." American Statistical Association 1979 Proceedings of the Survey Methodology Section, pp. 401-406.

sampling procedure also needed to be accounted for in generating the sampling weights.

The computation of the weighting component corresponding to the non-response followup was not an insignificant task. Numerous decisions, made after the nonresponse sample was selected, affected the computation of the sampling weight contribution. One such event was the return by mail of completed questionnaires from both sampled and nonsampled firms in the non-response sample, often as a result of the telephone prompting. This could have been handled most easily by ignoring any data that was returned by non-sampled nonrespondents and analyzing the questionnaires returned by the sampled nonrespondents in the usual manner. However, since the response rate was so low it was decided that data from all returned questionnaires would be utilized. This meant that after all the data had been received the non-respondent frame would have to be redefined to exclude any frame elements associated with completed questionnaires that were not obtained through the followup procedures.

Another event that necessitated corrections to the nonresponse frame was the return, by firms, of questionnaires that included data for more than one originally sampled firm. Unfortunately, this relationship was not identified in the questionnaire and was not confirmed until the telephone callbacks discussed in Section 4.3.2 were made to determine the level of aggregation reported in the data. Therefore, the frame of nonrespondents included some firms whose data had been reported by another sampling unit and hence, were not nonrespondents. This also had to be accounted for before computing the nonresponse sample component to the sampling weight.

The decision to collect total, national data for the industry giants also effected the frame of nonrespondents. The nonrespondent frame included some

elements associated with the three giants, and the data for such listings were not obtained by the nonresponse followup. The subsampling weight components were therefore adjusted to reflect this.

Procedures to compute the nonresponse subsampling weight components could not begin until all completed questionnaires had been returned and reporting aggregates identified. Those firms originally defined as nonrespondents and the corresponding firms that were selected for the nonresponse followup were so identified within each state and industry sector category. Then, any firm so identified that had not returned a questionnaire as a result of the follow-up procedures or was included in the aggregate of another completed questionnaire was flagged. The number of such flagged firms was then determined by state within industry sector. The number of nonrespondents and sampled nonrespondents within each state by sector stratum were adjusted downward by the number of corresponding firms that had responded independently of the followup procedures. The nonresponse subsampling weight component was then computed as follows:

$$1^{A}_{hij} = \begin{cases} 1 & \text{if response for firm-j was obtained without followup} \\ \frac{K_{hi}}{k_{hi}} & \text{if response for firm-j was obtained in followup} \\ 0 & \text{otherwise} \end{cases}$$

where

1^{A}_{hij} = the nonresponse subsampling weight component for firm-j from stratum-i in state-h

K_{hi} = the total number of actual nonrespondents in state-h and stratum-i after adjusting the original number of nonrespondents for questionnaires returned without the followup, firms that are a part of the industry giants, and firms included in another sample respondent's questionnaire

k_{hi} = the number of the K_{hi} nonrespondents that were sampled.

4.6 Determining Eligibility

4.6.1 Using Questionnaire Data

One major problem in constructing the state sampling frames was the inability to positively determine, using the state regulatory lists, whether or not a listed firm satisfied the criteria of the target population. For this reason the questionnaire was structured to help determine the eligibility status of a responding firm. In particular, question 3 provided the major input into eligibility determination. Respondents who reported that their firm did not furnish pest control services on a commercial basis were considered ineligible for the survey.

Answers to question 20 were also used to determine eligibility. Firms that reported their business to be 100 percent commercial agricultural, 100 percent pesticide sales or any combination of the two that totaled 100 percent of their gross sales were considered ineligible.

Other comments recorded in the questionnaire were also used to determine eligibility. Some comments that resulted in the firm being considered ineligible included the following: not in business in 1981; in business for only a very short time in 1981; only a very small percentage of the business was pest control services; and pesticides used only on the firm's property.

Another type of ineligibility was defined by firms that were out-of-business at the time of the survey. This information was provided either by the post office or by the person to whom the questionnaire had been forwarded.

4.6.2 Undeliverable/Unable to Contact Adjustment

When a sampled firm had no questionnaire data its eligibility could not be determined. A particular set of such firms were the undeliverables and the unable-to-contacts. These were firms that were mailed a questionnaire that was returned by the post office as undeliverable or firms in the nonresponse

followup that the interviewers were unable to contact despite repeated attempts. These firms could have been treated in various ways for weighting purposes. They could have been collapsed with the group of questionnaires that presumably reached the sampled firm but were not returned by that firm. Another option would have been to assume that all undeliverables and unable-to-contacts were associated with firms that were out-of-business and thus ineligible.

A third option, and the one which was chosen, was to determine if the firm was "in" or "out" of business. For the undeliverables this was done by calling directory assistance for a telephone listing under the name and address of the firm. If a telephone listing was obtained, the firm was considered to be "in" business. Otherwise, the firm was defined as "out" of business. All undeliverables from the second mailing of the questionnaire were included in this effort, whereas only a simple random sample of the first mailing undeliverables was included.

A slightly different procedure was used for the unable-to-contact firms. Prior to selecting the nonresponse followup sample, a telephone prompting was conducted for all firms that were in the nonresponse followup frame. The result of the telephone prompt was used to define the "in" or "out" of business status of firms sampled in the nonresponse followup that were classified as "unable-to-contact." The following defines the telephone prompting result codes:

<u>Code</u>	<u>Description</u>
0	Unknown status
1	No telephone number according to directory assistance
2	No contact after 3 attempts

- 3 Questionnaire returned
- 4 Intends to return questionnaire
- 5 Questionnaires not received but mailing address on label correct
- 6 Questionnaire not received and mailing address on label incorrect
- 7 Firm not prompted because questionnaire received in time
- 8 Firm received questionnaire and refused to respond
- 9A Questionnaire received but no other information
- 9B Prompting made but no response recorded by interviewer

Firms with codes of 1 or 2 were classified as "out" of business, while any of the remaining codes defined the firm as "in" business.

An indicator variable for "in" or "out" of business was set for each firm classified as undeliverable or unable-to-contact based on the procedures previously described. A special weight adjustment was also computed for the undeliverables that were sampled from the set of first mailing undeliverables. This adjustment was computed as follows:

$$2^{A_{hij}} = \begin{cases} 0 & \text{if firm-j from state-h and stratum-i was an undeliverable in the first mailing that was not selected for the in/out of business callback sample} \\ \frac{\sum_h \sum_i \sum_{j \in C} \frac{w_{hij}}{t_{hij}}}{\sum_h \sum_i \sum_{j \in C'} \frac{w_{hij}}{t_{hij}}} & \text{for firms in } C' \\ 1 & \text{otherwise, where} \end{cases}$$

w_{hij} = the sampling weight of firm-hij adjusted for the industry giants

t_{hij} = the estimated multiplicity factor for firm-hij (see section 4.9.1)

C = the set of all undeliverables from the first mailing

C' = the set of undeliverables from the first mailing
that were sampled.

4.7 Defining Respondents

Often, the major adjustment to sampling weights for surveys using probability sampling is the adjustment for nonresponse. For NUPAS, two general categories of nonresponse were recognized and the associated adjustments to the weights were calculated. The first general level of nonresponse occurred when all data was missing from the questionnaire. Included in this level were:

- Refusals
- Unreturned questionnaires
- Undeliverables and unable-to-contacts that were classified as "in" business.

The second level of nonresponse was defined by the data that was provided in question 22. Any firms providing information that had or could be used to generate an EPA registration number were defined as respondents.

A firm could be defined as a respondent even though their questionnaire had not been returned. This could result when one respondent's data included data for another sampled firm. It could also result when the data for a firm not selected in the sample was reported with data from a sampled firm. In both the above cases, data from the responding firm was duplicated and included on the data file for all sample members identified as belonging to the reporting aggregate.

4.8 Defining Nonresponse Weighting Classes

Weighting classes categorize the sampling units into what are expected to be homogeneous sets. Both responding and nonresponding sampling units are

assigned to a weighting class according to specific characteristics (exhibited by the sampling units) for the variables that define the weighting classes. Since all sampling units in a given weighting class have the same weighting class characteristics, it is expected that their responses to this survey would be similar. Therefore, nonresponse adjustments are computed within these weighting classes.

For NUPAS, the weighting classes used to adjust for nonresponse were defined by two variables, census region and industry sector category. The region variable could take on four values and was determined by the census region corresponding to the state where the sample member originally was listed on the sampling frame. This state was referred to as the reference state on the questionnaire.

The industry sector code assigned to a firm was determined whenever possible from data given in the questionnaire. If the firm was a non-respondent or did not provide appropriate and usable data to determine the sector category, then the default value was the sector code assigned in the sampling frame (see Technical Volume I, Chapter 2). In order of priority, questionnaire data that were used to determine the industry sector were: questions 21, 14, and 17; if needed, the types of products reported; and finally the firm name.

The steps to determine a selected firm's sector will be described in the remainder of this section. Any sampled firms that did not have associated questionnaire data either due to the lack of a direct response or due to not being included in another sampled unit's aggregate data were assigned to the industry sector corresponding to their sector code on the sampling frame. Since complete data for the industry giants were obtained, there was no need for sector categorization for weighting class purposes. For analysis

purposes, sector codes were assigned to the industry giants by company. Chemlawn/Chemscape were assigned to category 1 (tree and/or lawn), and Terminix and Orkin to category 2 (structural).

The sector was determined whenever possible by using data provided by the firm itself or the data provided for the firm by some other element selected from the sampling frame. The response to question 21 was the primary determinant of a firm's sector category. However, the percentages given in the response categories, excluding the "other" category, had to total at least 25 percent to be used to assign the sector category. If the ratio of the sum of a firm's percentages for the "Lawn and Turf" and "Tree, Shrub, and Ornamental" categories to the sum of its percentages for the "Termites and Other Wood Destroyers," "Commodity Fumigations," and "General Pest Control" was 10 to 1 or greater, then the firm was assigned to the industry category 1. If the inverse of that ratio was 10 to 1 or greater, then the firm was assigned to the industry category 2. If neither ratio was at least 10 to 1, then the firm was assigned to sector category 3 (tree/lawn and structural).

When the appropriate percentages in question 21 did not sum to at least 25 percent or when question 21 data were missing, question 14 was used to determine a firm's sector category. If the respondent indicated in question 14 that he/she or the firm was a member of only the National Arborist Association (NAA) or the Professional Lawn Care Association of America (PLCAA), then the firm was assigned sector category 1. If he/she or the firm were members of only the National Pest Control Association (NPCA), then the firm was assigned to sector category 2. If question 14 indicated membership in both NPCA and NAA or PLCAA, then the firm was assigned to sector category 3.

When it was reported in question 14 that the firm belonged to no national organization or if it belonged to something other than the NPCA, NAA, or

PLCAA, or if question 14 data were missing, then the original sector category on the sampling frame was checked. If the original sampling frame category was not 4 (unknown), then the firm retained its original sector category for the weighting classes. If the original sector category was 4, then the response to question 17 was examined for only those firms classified as eligible. If the respondent reported subscribing only to "American Lawn Applicator" or "Arborage" or "Grounds Maintenance" or "Landscape and Turf" or "Landscape Industry" or "Lawn Care Industry" or "Weeds, Trees and Turf", then the firm was assigned to sector category 1. If subscriptions were reported only to "Pest Control" or "Pest Control Technology" or "Pest Management", then the firm was assigned to sector category 2. If subscriptions were reported to magazines in both the category 1 and category 2 lists, then the firms was assigned to sector category 3.

If question 17 data were missing or not usable, then an examination of the products listed in question 22 along with the name of the firm were used to assign the sector category. If a firm reported strictly weed control and/or outdoor insect control products, then it was assigned sector category 1. If the products reported were indoor pest and termite control products, then the firm was assigned sector category 2. If a mixture of indoor and outdoor products were reported, then sector category 3 was assigned unless the firm name indicated it was either strictly a lawn and tree or a structural pest control business.

All eligible respondents that were classified on the sampling frame as sector category 4, were reclassified to a 1, 2, or 3 category using the procedures described above. Only four of the ineligible respondents having a sector category of 4 on the sampling frame could not be reclassified.

4.9 Nonresponse Adjustments

4.9.1 Multiplicity Adjustment for Nonrespondents

As discussed previously, the multiplicity for respondents was based on data provided. Unfortunately, for the nonrespondents such additional information was not available to make the same type of multiplicity determination. Based on the information about the respondents it was decided not to assign a multiplicity value of one to all nonrespondents, as would be done in a traditional nonresponse adjustment procedure. RTI used a more complex procedure whereby multiplicity was estimated for nonrespondents within each state by sector category. This estimate included contributions from both the mail and telephone state multiplicities as appropriate.

Even for respondents, the value of the telephone state multiplicity was often a problem. Callbacks to resolve potential multiplicity problems were related to the nature of the firm. However, in telephone states, multiplicity was related to the number of state-certified applicators employed by a firm. Therefore, a firm in a telephone state that reported only one location and service to only one state presented a multiplicity problem when the firm did not report its number of certified applicators. In such cases the firm's telephone multiplicity had to be estimated. The procedure for estimating the telephone multiplicity used the available data in the state by industry category strata. First it was necessary to compute the average number of certified applicators reported per business location. Then, that average was applied to the number of locations reported by the firm with the missing telephone multiplicity. The average number of certified applicators was computed as:

$$\bar{C}_{hi} = \frac{\sum_j T_{hij}^w A_{hij} C_{hij}}{\sum_j T_{hij}^w A_{hij} L_{hij}}$$

where

T_{hij}^W = the basic weight for firm-j in telephone state-h and sector category-i

l_{hij}^A = the nonresponse subsampling weight component for firm-j in state-h and sector category-i

C_{hij} = the number of certified applicators reported by firm-hij

L_{hij} = the number of business locations reported by firm-hij

The estimated telephone multiplicity, $\hat{M}_{hij}$, for firm-j in state-h and industry category-i was then calculated as:

$$\hat{M}_{hij} = \bar{C}_{hi} * L_{hij}.$$

Computing $\hat{M}_{hij}$ then gave all respondents a multiplicity value corresponding to the firm's service area among mail states and telephone states. Each firm's total multiplicity was defined as the sum of its mail state and telephone state multiplicity components.

With the total multiplicity for all respondents now established, an estimate of total multiplicity for nonrespondents could be computed as follows:

$$\bar{B}_{hij} = \frac{\sum_{j \in R} W_{hij} l_{hij}^A}{\sum_{j \in R} W_{hij} l_{hij}^A B_{hij}}$$

where $\bar{B}_{hi}$ = average, weighted total multiplicity for state-h and sector category-i

W_{hij} = the basic weight adjusted for the industry giants for firm-j in state-h and sector category-i

l_{hij}^A = the nonresponse subsampling weight component for firm-j in state-h and sector category-i

B_{hij} = the total multiplicity for firm-hij

R = the set of responding firms.

4.9.2 Adjustment for Refusals

Refusals were considered to be a special case of nonrespondents and thus were treated separately when computing nonresponse adjustments. Since refusals went to the trouble to respond to let us know they would not complete the questionnaire, we assumed the respondent considered the firm eligible and was unwilling to take the time to complete the questionnaire. Thus, all refusals were treated as eligible firms. This created a problem with all weighting classes defined by sector category 4 (unknown). There were refusals in that category but all eligible respondents originally in sector category 4 had had their sector category revised so that no eligible respondents remained in that category to assume the weight of the refusals in category 4. Therefore, the sum of weights for refusals in industry sector 4 was reapportioned to categories 1, 2 and 3 in proportion to the sum of weights for eligible respondents that had been reclassified from category 4 to categories 1, 2, and 3. This can be expressed as:

$$F_{hi} = \frac{\sum_{j' \in J'_i} \frac{w_{hij'} l_{hij'}}{B_{hij'}} \sum_{k \in R} \frac{w_{h4k} l_{h4k}}{\bar{B}_{h4k}}}{\sum_1 \sum_{j \in J} \frac{w_{hij} l_{hij}}{B_{hij}}}$$

where

$w_{hij'}$ = the basic weight adjusted for industry giants for firm- j' in census region- h and sector category- i

J'_i = the set of eligible respondents classified on the sampling frame as sector category-4 and reclassified as sector category-1

w_{h4k} = the basic weight adjusted for industry giants for refusal firm- k in census region- h and sector category-4

R = the set of refusal firms

- $1^{A_{hij'}}$ = the nonresponse subsampling weight component for firm-hij'
 $1^{A_{h4k}}$ = the nonresponse subsampling weight component for refusal firm-h4k
 $B_{hij'}$ = the total multiplicity factor for firm-hij'
 $\bar{B}_{h4k}$ = the average total multiplicity factor for refusal firm-h4k
 W_{hij} = the basic weight adjusted for industry giants for firm-j in census region-h and sector category-i
 J = the set of eligible respondents classified on the sampling frame as sector category-4
 $1^{A_{hij}}$ = the nonresponse subsampling weight component for firm-hij
 B_{hij} = the total multiplicity factor for firm-hij

Then the adjustment to the eligible respondents sampling weight can be expressed as:

$$3^{A_{hi}} = \begin{cases} 1, & \text{if firm-hij is ineligible} \\ 0, & \text{if firm-hij is a refusal} \\ \frac{\sum_{j \in E} \frac{W_{hij} 1^{A_{hij}}}{B_{hij}} + \sum_{j' \in R} \frac{W_{hij'} 1^{A_{hij'}}}{\bar{B}_{hij'}} + F_{hi}}{\sum_{j \in E} \frac{W_{hij} 1^{A_{hij}}}{B_{hij}}}, & \end{cases}$$

If firm-hij is an eligible respondent

where

- W_{hij} = basic weight adjusted for industry giants for firm-j in census region-h and sector category-i
 E = the set of eligible respondents
 $1^{A_{hij}}$ = nonresponse subsampling weight component for firm-hij
 B_{hij} = the total multiplicity factor for firm-hij
 $W_{hij'}$ = basic weight adjusted for industry giants for firm-hij'
 R = the set of refusal firms

$1^{A_{hij}}$ = the nonresponse subsampling weight component for firm-hij'

$\bar{B}_{hij}$ = the average multiplicity factor for firm-hij'

F_{hi} = the sum of weights for refusals in sector category-4 that were reassigned to sector category-i in census region-h

4.9.3 Weighting Class Nonresponse Adjustment

The final nonresponse adjustments were computed within census region by sector category weighting classes. The sector categories for the respondents were those reclassified using respondent data as discussed previously. Industry categories for the nonrespondents remained as defined in the sampling frame. The nonresponse adjustments were calculated by a ratio estimation procedure using the sampling weights in conjunction with all previous weight adjustments. The following equation presents the process used to compute the nonresponse adjustment for a given weighting class-hi:

$$4^{A_{hi}} = \frac{\sum_{j \in R} \frac{w_{hij} 1^{A_{hij}} 3^{A_{hi}}}{B_{hij}} + \sum_{j \in I} \frac{w_{hij} 1^{A_{hij}}}{B_{hij}} + \sum_{j \in U} \frac{w_{hij} 1^{A_{hij}} 2^{A_{hij}}}{\bar{B}_{hij}}}{\sum_{j \in R} \frac{w_{hij} 1^{A_{hij}} 3^{A_{hi}}}{B_{hi}} + \sum_{j \in I} \frac{w_{hij} 1^{A_{hij}}}{B_{hij}}}$$

where

w_{hij} = the basic weight adjusted for the industry giants for firm-j in weighting class-hi

$1^{A_{hij}}$ = the nonresponse subsampling weight component for firm-j in weighting class-hi

$2^{A_{hij}}$ = The undeliverable adjustment for firm-hij $\in u$

$3^{A_{hi}}$ = the refusal adjustment for weighting class-hi

B_{hij} = the multiplicity factor for firm-j in weighting class-hi

R = the set of eligible respondents

I = the set of ineligible respondents

U = the set of nonrespondents (i.e. unreturned questionnaires, undeliverables and unable to contacts) that were classified as "in" business.

The previous adjustment was computed for eligible respondents where a firm was classified as a respondent whenever any questionnaire data were given. An alternative definition for respondents was also developed and a corresponding nonresponse adjustment computed. The definition was based on the presence or absence of usable data in question 22 of the questionnaire which was the reported usage of pesticides by the firm. The nonresponse adjustment was computed using the same formula shown for $4A_{hi}$ except that the definition of R and U were revised. This nonresponse adjustment will be called $4A_{hi}^*$ and will be defined exactly like $4A_{hi}$ with the following changes:

- substitute R^* for R where

R^* = the set of eligible firms with usable question 22 data

- substitute U^* for U where

U^* = the set of nonrespondents classified as "in" business where nonrespondents includes these in U as well as eligible firms that did not provide usable data for question 22.

The adjusted sampling weight to be used in the analysis of NUPAS data can take two forms depending on the data being analyzed. Analyses of any data except the usage data in question 22 should use the weight in the following form:

$$Q_{hij} = \frac{w_{hij} \cdot 1A_{hij} \cdot 3A_{hi} \cdot 4A_{hi}}{B_{hij}}$$

For analysis of the usage data the weight expressed as follows should be used:

$$Q_{hij}^* = \frac{w_{hij} \cdot 1A_{hij} \cdot 3A_{hi} \cdot 4A_{hi}^*}{B_{hij}}$$

It should be noted that Q_{hij}^* could be used to analyze any of the data as long as only data corresponding to firms with usable question 22 data are included

in such analyses. This, essentially, ignores good data provided by firms that did not have question 22 data and makes a nonresponse adjustment for that data.

5. VARIANCE ESTIMATION

5.1 Variance Estimation in Complex Surveys

To document the precision of survey statistics, confidence intervals are typically quoted for sample estimates of population parameters. Such intervals are designed so that over a conceptual sequence of replicated samples surveyed under the same general conditions, the associated sequence of intervals will in the long run contain the true population value a specified percentage γ of the time (e.g. $\gamma = 95\%$). The validity of reported confidence coefficients (γ) depends in large measure on how good an estimate of the standard error or variance one uses in constructing the interval.

The variance of a survey statistic measures the average squared difference between the sample estimate and its mean value averaged over the conceptual sequence of replicated sample. Armed with an estimate of variance, a 95 percent confidence interval is typically formed as follows

$$I(\hat{\theta}) = \hat{\theta} \pm 2 [\text{var}(\hat{\theta})]^{1/2}$$

where $\hat{\theta}$ depicts a generic sample statistic and $\text{var}(\hat{\theta})$ denotes the variance estimate for $\hat{\theta}$. Estimation of variances for survey statistics is treated in numerous texts on probability sampling methods. These texts present variance estimation formulae for many of the probability sample designs used in practice, and provide the theory required for deriving formulae for special purpose designs like the NUPAS sample. These sample design specific variance formulae take explicit account of the precision effects of various design features.

The design feature that contributes most of the complexity to NUPAS variance estimation formulae is the multi-phase structure. For the seven

telephone states a three-phase sample design was executed with an initial stratified random sample subjected to telephone screening. At the second phase, eligible applicator/firm units identified during the phase I telephone screening effort were subsampled and subjected to a mail survey. In the third phase, a subsample of mail survey nonrespondents was subjected to personal interview follow-up. In the 'Mail State' portion of the survey, the initial telephone screening phase was eliminated leaving a two-phase design.

In addition to the multi-phase stratified sample design, two features of the mail and telephone sampling frames substantially complicate NUPAS variance estimation. The first of these features relates to the presence of substantial numbers of ineligible units on the sample lists. Some of this frame inefficiency resulted from the exclusion of so-called national giants and the Louisiana supplemental list members from the initial sample. These sample exclusions were supported by a special census in the case of "giants" and a supplemental list sample in the Louisiana case.

The second source of frame-related complexity results from the fact that multiple frame members can and do provide some or all of the data required for the same aggregate reporting unit. The number of frame units that link to the same aggregate reporting unit is referred to as the frame multiplicity. To obtain reasonably valid confidence intervals, these multiplicities must be ascertained and employed carefully in both the sample weighting and variance estimation process.

There exist several general purpose computer programs that calculate variances for complex survey statistics. These procedures assume that the sample can be successfully modeled as consisting of two or more independent replicated samples drawn from each primary stratum. For these general

purpose methods to work well, all subsequent stages or phases of sampling should be nested within these initial replicates, and the first-stage sampling rates should be negligible or reasonably equal across all primary strata. While it would have been possible to force the NUPAS design into this mold by specifying random half-samples drawn from the initial selections in each NUPAS primary stratum and then linking all subsequent phase participants to these pseudo-replicates, the large and variable NUPAS primary sampling fractions argued against this general purpose methodology. The next section outlines the form of the special purpose multi-phase variance formula developed for NUPAS.

5.2 Variance Estimation Formula

The purpose of this section is to illustrate the form of the special purpose variance estimator derived for NUPAS survey statistics. The goal was to provide reasonably unbiased¹ and efficient computation formulae tailored specifically to the multi-phase nature of the NUPAS design. Particular attention was paid to correctly handling the substantial degree of frame unit ineligibility² in a computationally efficient manner.

The three-phase contribution to variance associated with the seven telephone states will be depicted in terms of the phase I state by industry category strata labeled $i = 1, \dots, 14$. The second-phase industry type post-strata are nested within primary strata and are labeled by j , where typically j ranges from 1 through 3 or 4 depending on the presence of the

¹Assuming that nonsampling error did not exist, or, if it did, that it acted independently among observations and without directional bias. Given the quality of the raw data, the complexity of the analysis, and the level of nonresponse, the accuracy of such an assumption is questionable.

²Ineligibility here is determined with respect to the redefined (e.g. not including industry giant members), rather than the original, sampling structure.

fourth "unknown" industry classification in addition to 1 = tree/lawn, 2 = structural, 3 = both. The third-phase stratum index k is nested within the first two-phase strata and assumes two levels with 1 = mail survey respondents and 2 = mail survey nonrespondents.

For variance estimation purposes, the inverse probability and non-response adjusted sampling weights depicted by w_{ijkl} , for eligible survey participant l from phase III stratum ijk , will not be divided by the associated frame multiplicity M_{ijkl} as was the case in the initial weight formulation described in Chapter 4. For proper variance estimation, one divides M into the survey outcome variable, say y_{ijkl} , forming the transformed measurement $z_{ijkl} = (y_{ijkl}/M_{ijkl})$. The estimated eligible population total for the variate y can now be specified as

$$t = \sum_{i=1}^{14} \sum_{j=1}^{J_i} \sum_{k=1}^2 T_{ijk},$$

with T_{ijl} denoting the weighted sum of transformed z_{ijkl} values. With a_{ijk} denoting the number of eligible participants from phase three stratum ijk , the associated stratum total has the form

$$T_{ijk} = \sum_{l=1}^{a_{ijk}} w_{ijkl} z_{ijkl}.$$

In addition to these phase III eligible participant totals, the variance formula requires the corresponding weighted sums of squares

$$SS_{ijk} = \sum_{l=1}^{a_{ijk}} w_{ijkl}^2 z_{ijkl}^2.$$

The aggregates of these weighted totals and sums of squares over the phase III nonresponse follow-up strata $k = 1$ and 2 are depicted by T_{ij+} and SS_{ij+} respectively. The corresponding sums across the phase II industry type strata are denoted by T_{i++} and SS_{i++} . The overall sums added across the phase I state by industry class strata are T_{+++} and SS_{+++} .

The variance formula for the telephone state eligible population total t also requires estimated eligibility rates ϕ_{ijk}^3 , ϕ_{ij}^2 , and ϕ_i^1 for the three-phase specific levels of the design. If one operates on the full sample data set including the ineligible cases discovered at each phase, then the eligibility rate estimators are simply the weighted fractions of eligible cases accumulated to the design level specified by the stratum indicators. Combining these eligibility rate estimators with the T and SS totals, three sets of phase level mean squares are estimated as follows:

$$S3_{ijk}^2 = (w_{ijk+}/\phi_{ijk}^3) SS_{ijk} - T_{ijk}^2$$

$$S2_{ij}^2 = (w_{ij++}/\phi_{ij}^2) SS_{ij+} - T_{ij+}^2$$

and

$$S1_i^2 = (w_{i+++}/\phi_i^1) SS_{i++} - T_{i++}^2 .$$

The w_{ijk+} , w_{ij++} , and w_{i+++} quantities in the mean square defining equations above are simply the weight sums obtained by adding across the subscripts replaced by a 'plus' sign.

The variance estimator for the telephone state total t is now constructed from the $S1^2$, $S2^2$, and $S3^2$ phase specific mean squares as follows:

$$\begin{aligned} \text{var}(t) &= \sum_{i=1}^{14} \alpha 1_i S1_i^2 \\ &+ \sum_{i=1}^{14} \sum_{j=1}^{J_i} \beta 1_i \alpha 2_{ij} S2_{ij}^2 \\ &+ \sum_{i=1}^{14} \sum_{j=1}^{J_i} \beta 1_i \beta 2_{ij} \alpha 3_{ij2} S3_{ij2}^2 . \end{aligned}$$

The α coefficients in the variance estimator involve the finite population correction (fpc) factors for the associated levels of the design. For example,

$$\alpha 1_i = (1 - n_i/N_i) \div (n_i - 1)$$

is the first phase α coefficient with N_i indicating the phase I frame count from stratum i and n_i depicting the corresponding phase I sample size. Since the sampling rate from the phase III mail response post-stratum ($k=1$) is conceptually viewed as 100 percent, the associated $\alpha 3_{ij1}$ fpc related coefficient is zero. The corresponding terms involving $S3_{ij1}^2$ are therefore deleted by virtue of having a zero α coefficient.

The β coefficients provide unbiased variance estimation in the presence of small samples. For sufficiently large samples at each phase these β coefficients tend to one. For example,

$$\beta 2_{ij} = [m_{ij}/(m_{ij} - 1)] [(n_{ij} - 1)/n_{ij}]$$

where the n_{ij} denote phase I sample sizes distributed by phase II post-strata and the m_{ij} denote phase II subsample sizes. If the m_{ij} are large enough so that the first bracketed quantity in $\beta 2_{ij}$ is approximately one, then the $\beta 2_{ij}$ and $\beta 1_i$ coefficients will all be close to one and the variance formula reduces to the simplified form

$$\begin{aligned} \text{var}(t) = & \sum_{i=1}^{14} (1 - f1_i) S1_i^2/(n_i - 1) \\ & + \sum_{i=1}^{14} \sum_{j=1}^J (1 - f2_{ij}) S2_{ij}^2/(m_{ij} - 1) \\ & + \sum_{i=1}^{14} \sum_{j=1}^J (1 - f3_{ij2}) S3_{ij2}^2/(r_{ij2} - 1) \end{aligned}$$

where r_{ij2} depicts the phase III nonresponse follow-up sample size from phase II post-stratum ij . The $f1$, $f2$, and $f3$ quantities are the respective sampling rates for the associated phase I, II, and III strata. The NUPAS sample sizes are in-fact rather small for some phase strata and the associated β coefficients are nontrivially greater than 1.

The mail state variance contribution for an overall NUPAS population total is computed using a two-phase version of the telephone state formula. The variance of a combined NUPAS population total is then obtained by adding the mail state and telephone state contributions. A general description of the software developed for NUPAS variance estimation is provided in the following section.

5.3 Variance Estimation Software

RTI has implemented a computer program, using the SAS MACRO language, for computing estimated totals and estimates of variances. This software uses the formulae described in section 5.2. Estimates can be generated within user-specified domains as well as for the entire population.

The program logic is depicted in Figure 5-1. An initial pass of the data multiplies each of the analysis variables by the multiplicity/nonresponse adjustment factor and splits it into two SAS working files, one for the mail states and the other for the telephone states. Next, the SAS procedure PROC SUMMARY is used to generate, by domain, the first-, second-, and third-stage weighted sums and sums of squares used in the formulae. The dataset created by PROC SUMMARY is then processed to compute the estimate of the variance. Finally, the mail and telephone components are summed to create the estimate.

The output dataset contains the domain variables, estimates of the totals for each of the analysis variables, and the estimate of the variance for each of these totals.

RTI performed hand calculations on a subset of the data to verify the correctness of the software. Both the mail and telephone state variance component computations were checked at the subdomain and population levels. As with any complex, newly developed software, the possibility of underated errors exists despite such checks.

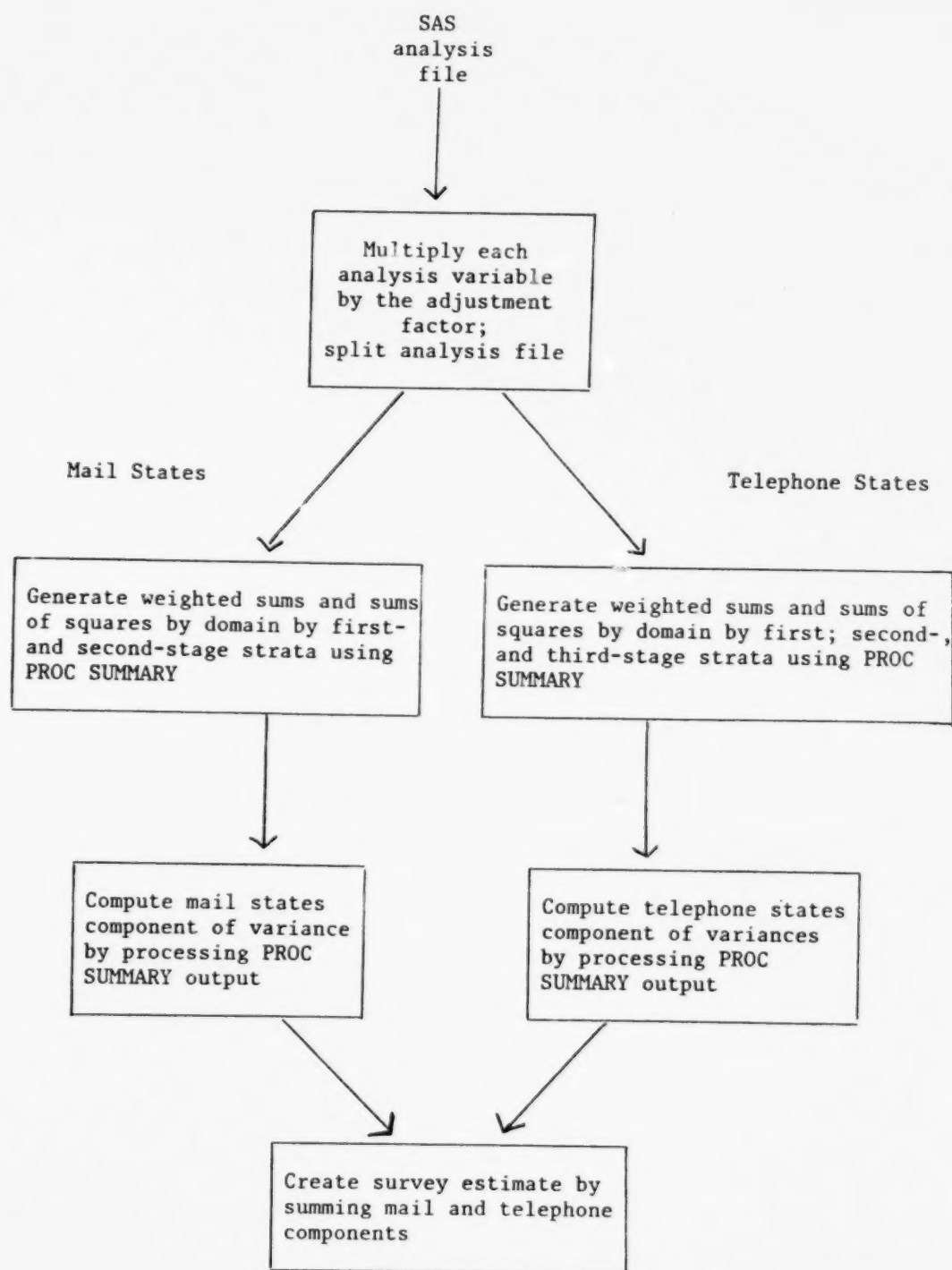


Figure 5-1. Flow of Logic for the Variance Estimation Software

5.4 Construction of the Usage Analysis File

To calculate estimates of variances for the total usage data using the software described in Section 5.3, the input dataset must have a record containing the following data for each sample member:

- stratum identifiers for each stage of sampling
- stratum population and sample size and eligibility ratios for each stage of sampling
- the raw weight
- an adjustment factor for multiplicity and nonresponse
- domain identifiers, including an eligibility indicator, and
- analysis variables.

The analysis variables are created specifically for each analysis, and represent the unweighted amount of active ingredient(s) (by product type and/or formulation) used by the sample member overall or in the geographic area (i.e., EPA region) of interest.

Cost and data quality considerations led to the decision to calculate variances only for the usage totals. Should estimates of variance for other totals be desired, they can be generated by RTI subsequent to issuance of the final report. It should be noted that the NUPAS user file will not contain the information necessary to correctly calculate variance estimates.

6. USAGE DATA CLEANING AND EDITING

6.1 Intended Utilization of Product Usage Data

The usage data reported in the questionnaire was needed to estimate the quantities of active ingredients that were used. Since the questionnaire asked PCOs to report their pesticide usage at the product level, it was necessary to translate this data to the level of active ingredients. The available resource of active ingredient information was a tape of the EPA Product Label File. RTI put this file, as well as other EPA pesticide data files, on-line in a SAS format to facilitate access to the information. In this data base, products are linked to active ingredient percentages by an EPA registration number. In order to summarize usage data to the active ingredient level, it was necessary to link each reported product to a valid EPA registration number. Since this was not consistently done by the PCOs in reporting their product usage, it became the focus of the data cleaning and editing process. Accurate assessment of the resulting data quality is dependent upon a detailed understanding of the steps involved in this process.

The primary effort of the process was to summarize the product data to a 'generic' formulation¹ level so that each unique formulation of a product could be assigned to a unique EPA registration number. In reality, any given product name can be associated with multiple valid EPA registration numbers. This is the case primarily for two reasons:

- (1) Different manufacturers produce 'different' products which are identical in formulation and often name as well, and

¹Formulation here connotes both the physical characteristics (e.g. wettable powder, emulsifiable concentrate) of the product and the percentage(s) of active ingredient(s) in the product.

- (2) One manufacturer can produce the same formulation under different names and registration numbers.

Since the final level of summarization was to be at the level of active ingredients rather than products, selection of a unique registration number to represent a category of products with identical formulations eliminated the ambiguity problem. The accuracy of translation to the active ingredient level could still be maintained and data reduction facilitated by minimizing the number of unique formulation entities.

6.2 Characteristics of the Raw Usage Data

The extent of cleaning and editing required was determined by the condition of the raw usage data. Ideally, every reported product would be identified by a valid and properly formatted EPA registration number which could then be computer-matched to the corresponding active ingredient percentages in the Product Label File. However, the usage data, as reported, was often incomplete or insufficient to perform an accurate reduction to the active ingredient level.

The raw usage data was incomplete in part due to the wording in the questionnaire: PCOs may have misinterpreted the terminology that was used. The importance of reporting an EPA registration number for a product was not adequately emphasized, and many respondents perceived that the extra effort that they would have to expend to report the registration number was not necessary. In addition, no explanation of how the number should be reported was included. An EPA registration number is normally displayed on a label with its component parts (i.e., a 6-digit company number and a 5-digit product number) hyphenated and with leading zeros dropped. Other designations are usually added on the front and back. With no explanation

as to which parts of the number on the label are significant, respondents interpreted the registration number in a variety of ways-- many of which were not transformable into a number that could be used. The requests in the questionnaire for 'complete product name' and 'formulation' were not sufficiently self-explanatory and were, therefore, also variedly interpreted. As a result, these entries were often insufficient to yield a correct EPA registration number through automated means. The result, in terms of cleaning and editing the data, was that each field of each record of usage data had to be referenced manually in order to make an accurate assignment of name and EPA registration number. Without a standard format or even a clear definition of the meaning of each field, any attempt at computerized matching between fields and the EPA pesticide data bases would be extremely inefficient. Not a single record of raw data had an EPA registration number in a format that was suitable for computerized matching without modification.

Beyond the identity of the products, complete and accurate reporting of quantities was essential for making reasonable estimates of active ingredient usage. Approximately 86 percent of the usage records had both amounts and usable unit specifications (e.g., pounds, gallons, etc.) reported. The accuracy of these reported amounts could not be quantitatively assessed in a survey of this type. The remaining 14 percent of the raw usage data was not reported with usable amount information. For approximately 8 percent, no amounts of product used were even reported. The other 6 percent consisted of unit specification problems. In this latter group, some were resolved after further scrutiny of the actual questionnaire. In these cases, resolution of units took more mathematical manipulation than the questionnaire editors were trained to handle. The

remaining unit specification problems involved the absence of specified units or units specified in overly general terms such as 'boxes', 'crates', 'tubes', etc. Some of these were resolved by contacting manufacturers regarding their packaging methods (see section 6.4). In the final analysis, half of the problems due to unit specification remained unresolved.

The raw usage data file consisted of nearly 25,000 records, each one representing the usage of a particular product by a particular respondent. Because extensive manual editing was necessary to prepare the raw usage data for analysis, the cleaning and editing task proved to be considerably more difficult and time consuming than originally anticipated.

6.3 Assignment and Editing of Registration Numbers

Under the original plan, EPA was to edit the usage data. The condition of the raw data and the need to assess all information on each of nearly 25,000 records, made this an unworkable plan. Because of its better access to hard copy questionnaires and to the raw data file on the computer, and familiarity with the structure of the data file, RTI was reassigned responsibility for usage data editing.

The first step in the editing process was to define small and workable subsets of the raw data file for the purpose of assigning valid EPA registration numbers. Subsets were selected by taking groups of records containing similar product names. Within these subsets, like formulations of a product were identified and assigned a unique name. A unique EPA registration number was then assigned to each of these product groups by using the number most frequently referenced. In many records the reported formulation included information on the percentages of active ingredients. The reported formulation in the records only described the physical form of the product.

To aid in the assignment task, an 'edit flag' was added to each usage data record to reflect the level of editing performed on that record. Table 6-1 illustrates the meaning of the flag variable. As subsets of the raw data were edited, a list of assigned names and corresponding EPA registration numbers was also maintained and used as reference for the editing of subsequent subsets. This helped to maintain consistency throughout the editing process.

Once this process was completed for all records, the original list of 6,304 alphabetically unique product names was reduced to 1,350. Of these, 991 had associated EPA registration numbers (includes edit flag values 5, 4 and 2), 142 were names with no associated registration number (edit flag value of 3), and 217 were either unidentifiable or not pesticide products (edit flag value of 1).

Once the registration numbers were correctly formatted, the computer became more valuable in the editing process. To help verify what had already been done and to derive additional information for further editing, the usage data with EPA registration numbers was compared against EPA pesticide data bases. Because these data bases are not uniformly maintained, matches were not always found throughout all the data bases. Of particular interest was, of course, the Product Label File, as this data base contains the active ingredient percentages needed to reduce the usage data. Another file of interest for editing purposes was the Product Name Information File which linked EPA registration numbers to product names. Of the 991 unique registration numbers identified in the usage data, 941 were found in the Product Name Information File. Of that 941, all except about 20 were found in the Product Label File. The usage data was further edited to reflect information gained through the data bases and edit flags

Table 6-1. Initial Meanings for Edit Flag Values

Value	Initial Meanings for Edit Flag Values
5	Large number of references made to given product formulation and EPA reg. no.
4	Product formulation and EPA reg. no. appeared logical and valid, but few references made to combination.
3	Product name appeared to refer to a pesticide product, but no reference to an EPA reg. no. found.
2	EPA reg. no. looked legitimate, but information in product and formulation field inadequate to assign generic name.
1	Information in records inadequate to make any assumptions on identity; product referenced not a pesticide product.

were recoded to reflect the new editing status. This recoding included the addition of a new edit flag value as shown in Table 6-2.

The final step of the editing process was an attempt to resolve any usage data which had not yet been successfully linked to active ingredient information. Acceptable alternative EPA registration numbers were found for many cases by further searching of the EPA pesticide data bases. The remaining cases were resolved through consultation with pesticide industry experts. When completed, the original list of 6,304 unique names was reduced to 1,318. This includes 1,015 products for which active ingredient percentages are known, 33 that were identified in the Product Name Information File, but are not traceable to the active ingredient level, and 270 which are either unidentifiable or not pesticide products.

In terms of total records for usage data, the final status is best shown in Figures 6-1 and 6-2. The 93.68 percent "REDUCIBLE TO AI" shown in Figure 6-1 requires some qualification as it includes records with uncertain quantities and unit specifications as discussed in Section 6.2 of this document. Figure 6-2 breaks down this 93.68 percent to reflect this situation.

6.4 Resolution of Nonspecific Unit Codes

Following the assignment of EPA registration numbers, RTI attempted to translate the nonspecific unit information (coded as "L" by the hardcopy editors) into units of pounds or gallons. RTI's effort focused on those products that were assigned valid EPA registration numbers.

The usage file was subset to the 1,200 records that had valid EPA registration numbers and the unit code "L". Hard copy was examined for each of these records, and the nonspecific unit was categorized using the codes presented in Table 6-3. The range of updated unit codes and re-

Table 6-2. Revised Meanings for Edit Flag Values

6	EPA reg. no. found in Product Label File, active ingredient percentages known.
5	EPA reg. no. found only in Product Name Information File; reg. no. valid but active ingredient percentages not known.
4	EPA reg. no. not found in EPA pesticide data bases.
3, 2, & 1	As defined in Table 6-1.

USAGE DATA STATUS

NATIONAL URBAN PESTICIDE APPLICATOR SURVEY

100CT83

STATUS

REDUCIBLE TO AI

VERIFIED REGNO

UNIDENTIFIABLE

FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
23394	23394	93.68	93.68
95	23489	0.38	94.06
1484	24973	5.94	100.00

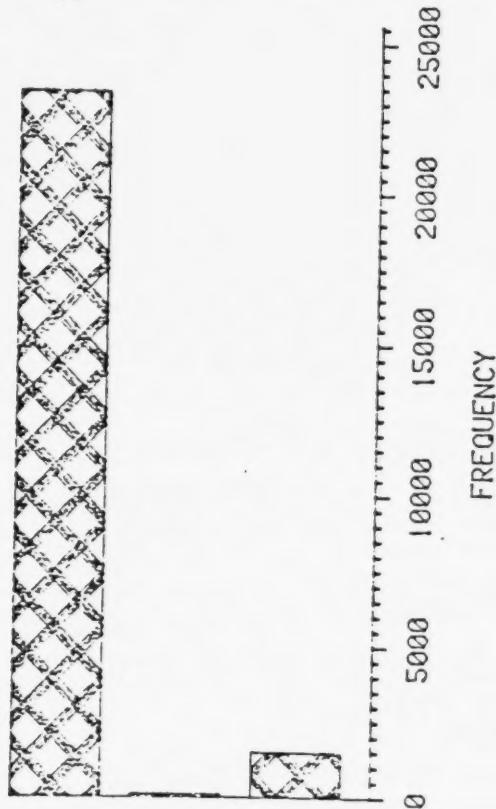


Figure 6-1

USAGE DATA STATUS NATIONAL URBAN PESTICIDE APPLICATOR SURVEY

11OCT83

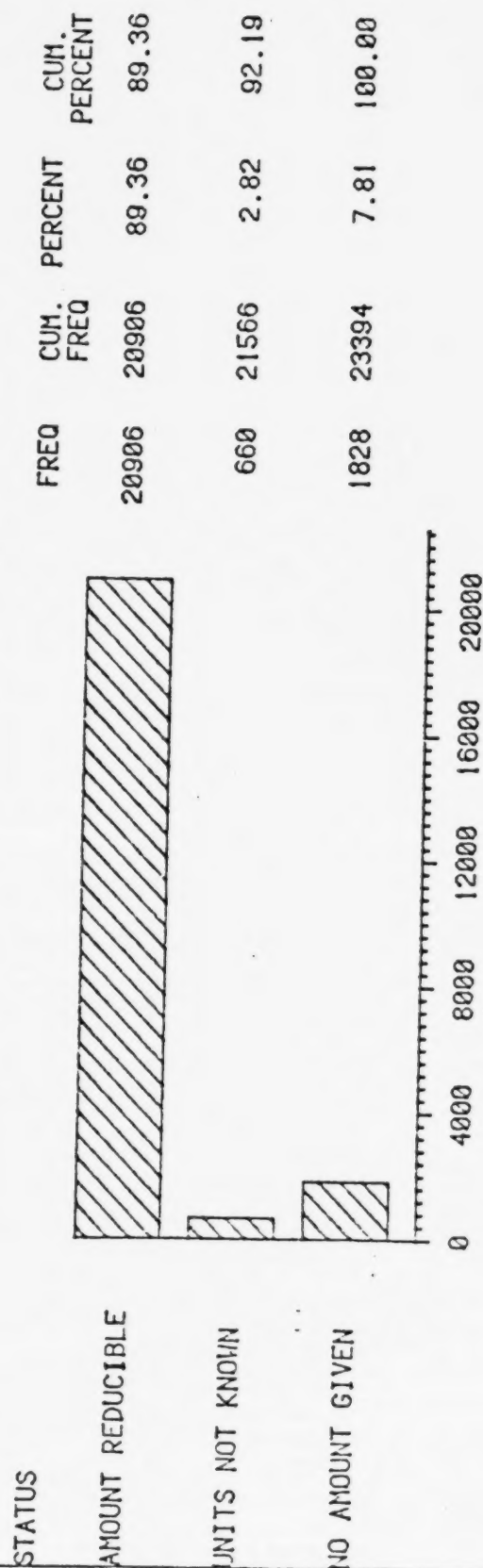


Figure 6-2

FOR DATA WITH KNOWN AI INFORMATION ONLY

Table 6-3. Unit Codes for Nonspecific Packaging Information

Unit Code	Description
M	cases
N	capsules
O	flasks
P	tubes
Q	cans
R	units, each, aerosols
S	packets, packs, packages
T	dozen
U	boxes, cartons
V	bottles
W	shots
X	bombs, cartridges
Y	bags
Z	drums
3	strips
4	cylinders
6	jars
7	pails
8	vials

sponses to the formulation section of question 22 was tabulated for each EPA registration number included in the above effort. The product manufacturer, identified through the EPA registration number, was then contacted and asked to provide packaging information that would permit transformation of the nonspecific unit.

In the majority of cases, the manufacturer could provide the desired information, which was then used to calculate the number of pounds or gallons of product reported. For some products, more than one size was possible for a unit code. In such instances, if one package size was sold more frequently than the other(s), it was used in all transformations; otherwise, the smallest package size was applied so as to be conservative. The potential introduction of measurement error by this process was recognized; however, it was considered to be smaller than the potential error associated with other treatments of these records. The cases that could not be resolved fell into four general categories.

1. The product was not packaged by the manufacturer in the units described. This was sometimes due to repackaging by distributors, or reflected unconventional packaging descriptions by the respondents.
2. The product name and formulation information provided by the respondent was not specific enough to define even a limited range of possible package sizes.
3. The product manufacturer could not be contacted, usually because it was no longer in business.
4. The manufacturer did not know of a product by the name reported on the questionnaire. This may have occasionally been due to misassignment of the EPA registration number, although in most cases it seems to have been due to renaming of the product by local distributors and formulators.

Of the 1,200 records examined, 873 were resolved and their amounts transformed into usable units. This resolution was particularly important for some products that had originally been reported primarily with non-specific unit codes.

6.5 Amount Imputation

A significant number of the 23,151 usage records that were assigned a valid EPA registration number and came from eligible respondents lacked either amount information (2,071 records) or usable unit codes (360 records after all possible nonspecific unit code transformations were done). The options for overcoming these deficiencies fell into two general classes: weighting class adjustments or item imputation.

Considerations of processing ease, schedule, and cost led to a decision to use a product-level imputation scheme to replace missing or otherwise unusable amounts at the record level. The methodology for item imputation was as follows.

1. The usage data file was subset to eligible respondents, and to those records with valid EPA registration numbers. The amount data in any records with a usable unit code was transformed (if necessary) into the number of pounds of product used. The transformation for units of liquid measure was based on the density of water (i.e. 1 gallon = 8.35 pounds) in order to be consistent with assumptions incorporated in the EPA pesticide data bases.
2. For each product, weighted averages of the amount used by firms of similar size within an EPA region were calculated (if possible). Four firm-size categories (\$0-\$49,999, \$50,000-\$199,999, \$200,000-\$999,999, and > \$1,000,000) were defined and filled based on response to the question on total gross sales (question 18). Weights adjusted for usage nonresponse (see section 4.9.3) were used in the calculations. The firm size by EPA region level of summarization was chosen because it provides some assurance that imputed values will be based on data from firms with similar characteristics, and yet is not so fine a level that many imputation classes would have no members with usable data. Weighted averages for each product were also calculated for firms of similar size across EPA regions, and for the product overall.
3. If an appropriate firm size by EPA region average had been calculated, it was used as the imputed value in all records for the product that did not have usable amount information and were from firms of similar size serving the same region. If the necessary size by region average had not been computed, the average for firms of similar size was imputed. Failing that, the product average was used. Of the 2,431 records which originally did not have usable amount information, 1,611 received imputed values at

the size by region level, 410 at the size level, and 357 at the product level. The remaining 53 records were for the 34 products which had never been reported in NUPAS with usable amount information.

Any technique used to overcome deficiencies in the data is based on certain assumptions which, if not met, introduce a potential source of error in the estimates computed from the data. The imputation method described above is no exception. It starts with the assumption that there is only minor variation in the amount of a product used by firms of similar size and in similar regions, regardless of whether or not they provided usable amount information for that product. More variation is expected to exist among firms of similar size operating in different areas of the country, and still more over the whole set of firms which use the product, but at each level the variation is still assumed to be relatively minor. Quantitative assessment of the impact of deviations from these assumptions has not been attempted. However, the accuracy of the final estimates of usage, which incorporated the imputed values, was undoubtedly better than the estimates that would have been produced without them.

6.6 Product Class and Formulation Categorization

All of the usage estimates generated in the NUPAS analyses are expressed as pounds of active ingredient, rather than product. However, it was necessary to retain some product-level attributes for the analyses. In particular, the product class (e.g. insecticide, herbicide, fungicide, etc.) and formulation (e.g. wettable powder, dust, emulsifiable concentrate, etc.) were required. These designations could not be made at the active ingredient level because some active ingredients occur in more than one class of product, and many of them occur in more than one formulation. Therefore, designations made at the product level were retained during conversion of product amount to pounds of active ingredient used.

Each reported product that was assigned a valid EPA registration number was assigned to a class and a formulation category based on variables in the RTI version of the computerized EPA Product Label File. For both characteristics, the variables on the file provided more specificity than was warranted, so they were mapped to a more general classification system. The procedure for doing this is described below.

Product class categories were defined using the type codes and product manager code reported on the Product Label file. Each product can have up to eight distinct type codes, some of which are not pertinent to the desired class categorization scheme. For those cases in which the pertinent type codes indicated that a product could belong to more than one class, the product manager number was used to decide which class to assign it to. The mapping from type codes and product manager codes to the final product class categories is shown in Table 6-4.

Formulation categories were defined directly from the formulation codes on the product label file. The mapping from formulation codes to categories is given in Table 6-5.

6.7 Estimation of Active Ingredient Usage

Once all products assigned valid EPA registration numbers were categorized by class and formulation and all imputations of product amount completed, the pounds of product in a record were converted into pounds of one or more active ingredients. The conversion was accomplished by using the "percent active ingredient" data on the RTI version of the EPA product label file. Active ingredients are identified on the file by a unique chemical code number assigned by EPA. All chemical codes associated with a product were retained, including those for solvents. Overall, the 975 valid EPA registration numbers which were reported for or assigned to

Table 6-4. Definition of NUPAS Product Class Categories

Product Class Category	Product Manager Codes ¹	Type Codes ²	Type Code Definition
Vertebrate Pesticide	-	2	Poison, multidose
		3	Poison, single dose
		4	Repellent
Microbicide	31, 32	22	Disinfectant/bacteriocide/germicide
		23	Sanitizer
		24	Bacteriostat
		25	Virucide
		28	Microbicide/microbistat
Fungicide	21	30	Fungicide and nematocide
		31	Fungicide
Herbicide	23, 25	40	Herbicide, unspecified
		41	Algaecide
		44	Plant growth regulator
		46	Herbicide, terrestrial
		47	Herbicide, aquatic
Insecticide	12, 15, 16, 17	32	Nematicide
		50	Insecticide and miticide
		51	Miticide
		53	Molluscicide
		54	Repellent or feeding depressant
		57	Pheromone, growth regulator, attractant
		59	Biological agents

¹These were used only if the type code mapping was not unique. Product manager codes for vertebrate pesticides were not used.

²Only valid, applicable type codes associated with products reported in NUPAS are shown.

Table 6-5. Definition of NUPAS Product Formulation Categories

Formulation Category	Formulation Code	Formulation Code Description
Dust	3	Dust
	7	Wettable powder/dust
Granular/pelleted	4	Granular
	5	Pelleted/tablet
Wettable powder	6	Wettable powder
Liquid concentrate	12	Emulsifiable concentrate
	14	Flowable concentrate
	15	Soluble concentrate
Ready-to-use solution	16	Solution, ready-to-use
Pressurized	18	Pressurized gas
	19	Pressurized liquid
	20	Pressurized solid
Other	1	Technical chemical
	2	Formulation intermediate
	8	Crystalline
	9	Microencapsulated
	10	Impregnated materials
	17	Oils-no added pesticides

products used by eligible NUPAS respondents contained 362 distinct active ingredients.

The unweighted amounts of active ingredients were then weighted using the usage nonresponse-adjusted weights described in section 4.9.3. Estimates of total usage in an analysis class were obtained by summing over all records in that class. Corresponding estimates of the variance of the totals were calculated using the variance estimation software (see sections 5.3 - 5.4).

7. EDITING NONUSAGE DATA FOR LOGICAL CONSISTENCY

7.1 Intended Level of Editing

Because the primary goal of NUPAS was to estimate the amount of pesticides used, the principal focus of the NUPAS data cleaning and editing effort was on the usage data. However, it was also desirable that the other (nonusage) data accurately represent the reporting aggregate that provided it. Schedule and resource considerations limited the nonusage editing to the most important data. Highest priority was placed on editing nonusage data that was needed for analysis weight calculation, usage imputation, and/or analyses involving usage. Other nonusage data was subjected to a reduced level of editing.

Although the hard copy edit process described in Chapter 2 improved data quality by standardizing responses and noting obvious respondent misinterpretations, further editing of many nonusage questions was required. This secondary editing examined the logical consistency of responses to multiple questions on a questionnaire. Examination across questions revealed and helped resolve three types of problems:

1. Miscodes and keying errors
2. Respondent misinterpretation of questions
3. Responses to different questions on the same questionnaire at different levels of aggregation (e.g. answers to some questions were for one location, whereas answers to other questions were for multiple locations).

The last type of problem was found with relative frequency during the reporting aggregate callback effort (see section 4.3); it was due in part to misinterpretation, and in part to the tendency to answer questions with readily available data, regardless of what aggregate it was for. During the callbacks, RTI placed primary emphasis on identifying which aggregate

the usage data represented. Time did not permit thorough checking with each respondent that all answers were for the same aggregate.

7.2 Editing of Specific Questions

Secondary nonusage data editing for logical consistency centered on six questions. For each, the editing process began by calculating a one-way frequency distribution of responses on the questionnaire file, or crosstabulation with responses to other questions. Anomalies were identified, and the original questionnaires examined. The general results of the editing are described below for each of these questions.

1. Number of firm employees (question 10). Two general forms of respondent misinterpretation of the question were found. In some very small firms, it was at times obvious that the owner did not consider himself/herself as an "employee", and so did not include himself/herself in the counts of employees. This led to a number of respondents (280) who reported "zero" employees in question 10. In such cases, the firm was logically imputed to have one technical employee. The other type of interpretation which led to anomalous response was found when the sum of parts a-f (number of technicians, managers, sales people, clerical staff, and other) was compared to part g (total number of employees) and found to exceed it. In 14 such cases, the respondent seems to have counted employees who fall in multiple categories more than once. Because the difference between the sum of a-f and the response to g was small in all cases, the reported numbers were used as is.
2. Number of certified/licensed employees (question 11). This data was critical for determination of multiplicity for respondents who served "telephone" states (see section 4.3). Editing of responses to this question was hampered by the variation among states in their regulatory programs, and by nonstandardized definitions of certification and licensing. Consideration was given to employing a consultant to obtain the data necessary for informed editing and interpretation of the responses to this questions. Appendix C of this volume describes the type of information required. Due to the perceived difficulty of obtaining much of this information, a subsequent decision was made not to pursue the matter further. A number of respondents reported zero certified/ licensed employees. In many cases, this may have been due to the first type of misinterpretation discussed above for question 20, while in others it may have been an accurate response. Those respondents who were selected in telephone states and reported zero certified/licensed employees were imputed to have one such employee since they were on a sampling frame constructed from lists of certified applicators. "Zero" responses for mail state sample members were left as is.

3. Amount of total gross sales (question 18). Responses to this question were cross-tabulated with responses to several other questions in an effort to check respondent interpretation and consistency of reporting level. Cross-tabs with the question on amount of sales from pest control services only (questions 19) identified some records for which question 19 was greater than question 18. Several of these were keying errors, and the rest appeared to be respondent error. In those cases, the response for question 19 was changed to equal the response to question 18. Crosstabs were also done with the total number of accounts reported in question 20 and with the number of business locations reported in question 5. Only a small number of anomalies (e.g. large number of accounts and small amount of gross sales) were identified. Of these, some were due to keying errors, some were attributable to respondent error or misinterpretation (e.g. a dollar value reported for number of accounts), and a few were interpreted as being legitimate representations of the reporting aggregate's business.
4. Sites served (question 20). As was discussed above, the "number of accounts" sections of this question were checked by cross-tabulating the responses with gross sales (question 18) responses. The percents reported in a questionnaire for this question were summed, and hard-copy checked for the 294 respondents who provided data which did not sum to 100%. Most (184) of the deviations from 100% were small and attributable to rounding error. The remaining deviations were more sizable, and most were due to respondent misinterpretation (e.g. reporting the percent of gross sales, rather than pest control services sales, associated with the sites) or partial response. A decision was made not to adjust the reported percents on the file so that they summed to 100%, but rather to adjust them during analysis if desired. In addition to the above checks, questionnaires were examined for the 43 respondents who reported 95% or more in the "other" section of the question (10g.). Based on notes written on the questionnaires, some of the responses were determined to be legitimate, but the majority reflected respondent misunderstanding of the question (e.g. outdoor pest control respondents interpreting the question as asking for the percent of business inside the defined sites, and so using the "other" category to indicate that their service was provided outdoors). In a few such cases, it was possible to correct the response based on other information on the questionnaire.
5. Industry sector served (question 21). The percents for this question were summed, and compared with the expected sum of 100%. For 109 of the 234 respondents who provided data which did not sum to 100%, the deviation was small and attributable to rounding error. In the other cases, the difference was due to partial response or respondent misinterpretation (e.g. reporting percent of gross sales, rather than sales of services, associated with the sectors). Responses were not adjusted on the file to sum to 100%. Hard copy was also examined for the 33 respondents who reported 95% or greater in the "other" category. Notes written

on the questionnaires indicated that some of these responses were accurate, and that others were due to respondent misunderstanding of the generality of the sectors listed (e.g. not wanting to include "weed control" in the lawn/turf and/or tree/shrub/ornamental categories). Some responses were altered based on other information in the questionnaire.

6. Time period the usage data was reported for (questions 23 and 24). Hard copy was examined for respondents who indicated that the usage data was provided for some period of time other than 1981. Reported period of less than 1 year were generally attributable to one of two causes.
 1. The respondent had been in business for less than 1 year.
 2. The respondent provided a seasonal service, and reported the months in which the service was provided.

Reported periods of greater than 1 year in most cases seem to reflect respondent error. Because the vast majority of respondents who reported periods of other than one year seem to have in actuality provided annual data, no adjustments of any data were made based on response to this question.

8. EXTERNAL REVIEW OF PRELIMINARY RESULTS

8.1 Motivation for External Review

At several points during data editing and analysis weight development, RTI generated preliminary estimates of pesticide usage and some nonusage parameters. These were provided to people not employed by RTI for review and comment. NUPAS was designed to be a cooperative effort between EPA, the trade associations, and the states, and review by members of the pest control industry was planned from the beginning. External (i.e., non-RTI) reviewers of preliminary NUPAS estimates included EPA personnel, staff and some members of industry trade and state associations, representatives of selected manufacturers of pesticides, and consultants regarded as experts on the nonfarm pest control industry (see Table 8-1).

The external review process was intended to independently assess the accuracy of the NUPAS estimates. The data cleaning and weighting procedures required many assumptions, and the complexity of the processes meant that there were many steps at which mistakes could be made. The only way to assess the accuracy of the assumptions and the logic of the procedures was to compare the estimates produced by them with certain benchmark estimates obtained by some other means. RTI personnel are not experts on the pest control industry, and for many of the necessary comparisons published, readily available benchmark estimates were not available. For this reason, the input by the external reviewers played a large role in the quality control aspects of NUPAS estimate generation.

8.2 Results of the Review Process

The review process resulted in major improvements in the quality of the estimates produced from the NUPAS data. Many of the preliminary

Table 8-1. Reviews and Presentations of Preliminary NUPAS Results

Reviewer or Presentation	Estimates Presented	Date
National Pest Control Association annual meeting	Coverage rates, industry descriptions, data quality descriptors	October, 1982
American Association of Pest Control Operators annual meeting	Coverage rates, industry descriptors	November, 1982
Union Carbide Agricultural Products, Inc.	Usage estimates for four active ingredients	June, 1983
Association of Structural Pest Control Regulatory Officials annual meeting	Usage and nonusage estimates	October, 1983
NPCA Project Development Council annual meeting	Usage and nonusage estimates	November, 1983
Orkin Pest Control, Terminix International, and Velsicol	Selected usage estimates	October - November, 1983
Interagency Pesticide Usage Planning Committee	Usage estimates	November, 1983

estimates were lower than comparable, independently derived estimates. The discrepancies led to a reevaluation of data editing and weighting techniques; they also resulted in improved techniques in data editing and the discovery of logic and/or programming errors in the software for usage imputation and the calculation of analysis weights. These changes resulted in final estimates that in many cases are substantially different from the preliminary ones.

It must be stressed that in no sense were the preliminary estimates adjusted to match external totals. The final estimates are based strictly on data gathered in NUPAS, and differences between the preliminary and the final estimates are due solely to improvements in data editing and analysis procedures.

9. DATA QUALITY

The quality of the estimates computed from the NUPAS data is a complex function of many factors, most of which have been described in the preceding chapters. The intent of this chapter is to describe the degree to which these factors are perceived to have affected the quality of the estimates, and to indicate the strengths and weaknesses of the NUPAS data.

9.1 Realized Target Population

Originally, the target population for this survey was to be all non-agricultural pest control firms serving the United States in 1981, excluding those which served only golf courses. Because state regulatory lists were used to construct the sampling frames, the target population had to be redefined; the realized target population included just those firms regulated by or employing people regulated by the 48 cooperating states and the District of Columbia. Quantitative estimates of the difference between the original and the redefined target population would be difficult to develop (see Appendix C), and have not been generated.¹

The sampling frame's coverage of the redefined target population also has not been quantified. Overcoverage, or the inclusion of firms not eligible for NUPAS, was greater than expected. Identification of some ineligibles was based directly on information which they provided and was straightforward. Elimination of the effect of other ineligibles was less

¹It should be noted that only 22 percent of responding reporting aggregates indicated that they used any restricted pesticides. Conceivably, this information in conjunction with information on state regulatory programs (e.g. was certification necessary for any commercial pesticide application, or only for application of restricted pesticides) could be used to assess the magnitude of the difference between the original and the target population.

direct, and relied on an assumption that the proportion of ineligibles to eligibles was similar in nonrespondents and respondents. Undercoverage, or the omission of members of the target population from the sampling frame, is thought to have occurred for the following reasons.

1. In some states, firms may have begun operations in a state in 1981 after the regulatory list provided by the state was last updated.
2. Some state lists did not include firms regulated through reciprocity agreements with other states.
3. Firms in specialized sectors (e.g. commodity fumigation) were not included on lists provided by some states.

Because the results of some NUPAS data analyses are expressed in terms of "number of target population units" with a particular trait, definition of "target population unit" is necessary. The definition of the ultimately realized units is not straightforward. The decision, prior to sample selection, to try to obtain state-level data from sample members meant that the target population unit was defined then as the state-specific portion of the activities of either a location or a firm, depending on the level at which the respondent chose to respond. Thus, a firm with multiple locations and/or which served multiple states could be represented by more than one unit. The fact that some respondents provided data for more than one frame element led to the final definition of target population units in terms of reporting aggregates. Reporting aggregates in this survey can represent any of the following;

- a state-specific portion of a single location's activities, or
- a state-specific portion of a firm's activities, or
- a single location, or
- an entire firm.

Therefore, any estimate that is expressed in terms of target population units is difficult to interpret in absolute terms.

9.2 Coverage Rates

The weighted coverage rate of a survey is the proportion of the weight ultimately attributed to the target population that was originally associated with the respondents. It can be defined for subsets of the target population, and at various stages of sampling. For NUPAS, coverage rates for each weighting class have been computed for the original sample that received the mail questionnaire, and for the sample following the non-response followup subsampling (see Tables 9-1 and 9-2 respectively). The coverage rates following subsampling for respondents who provided usage data are given in Table 9-3. The overall coverage rate following subsampling of 74.6 percent is relatively high for an initial mail survey of businesses.

9.3 Potential Sources of Error

The estimates calculated from any survey data have associated with them a number of potential sources of error. These sources can act to different degrees, and may act in concert with or in opposition to each other. Sampling errors are those that are strictly functions of the sample structure, sample size, and the realized values of the variables. In this survey, sampling errors are approximated using the formulae and software described in Chapter 5. All other sources of error are often collectively referred to as nonsampling error. Sampling and nonsampling errors are not independent, since sampling errors are functions of realized values, which are affected by nonsampling errors.

Considerations of data quality and response characteristics lead to a general but unquantified impression that nonsampling errors are of at least

Table 9-1. Coverage Rates for Eligible Respondents Prior to the
Nonresponse Followup Subsampling

Census Region	Industry Category			Overall
	Tree/Lawn	Structural	Both	
1	0.576	0.631	0.804	0.617
2	0.530	0.473	0.459	0.499
3	0.612	0.537	0.354	0.512
4	0.441	0.654	0.497	0.566
Overall	0.551	0.559	0.444	0.539

Table 9-2. Coverage Rates for Eligible Respondents After the
Nonresponse Followup Subsampling

Census Region	Industry Category			Overall
	Tree/Lawn	Structural	Both	
1	0.678	0.844	0.808	0.760
2	0.788	0.776	0.711	0.776
3	0.756	0.801	0.518	0.732
4	0.557	0.843	0.572	0.714
Overall	0.720	0.811	0.590	0.746

Table 9-3. Coverage Rates for Eligible Respondents Who Provided Usage Data,
After the Nonresponse Followup Subsampling

Census Region	Industry Category			Overall
	Tree/Lawn	Structural	Both	
1	0.543	0.710	0.757	0.632
2	0.677	0.604	0.571	0.636
3	0.633	0.658	0.371	0.592
4	0.457	0.633	0.463	0.553
Overall	0.600	0.652	0.461	0.605

the same order of magnitude as sampling errors in NUPAS. The following sources of nonsampling error had the potential to affect the final estimates.

1. Inaccurate mapping of reporting aggregates to the sampling frame and sample members. Due to variations and inaccuracies in firm name and address information on the state regulatory lists, some members of a reporting aggregate may not have been identified as such based on the telephone callback response. This would have resulted in a multiplicity factor for the reporting aggregate that was too small. This in turn would have affected the analysis weight. Inaccuracies in mapping could also have introduced double coverage if a respondent was not properly identified as belonging to another reporting aggregate. This source of error is not believed to be a major contributor to the nonsampling error.
2. Undercoverage and/or overcoverage of the target population. This was discussed in section 9.1. Undercoverage is known to have occurred, and acted to reduce the estimates. The impact of coverage error on the estimates is thought to vary among industry sectors and among active ingredients.
3. Incorrect assignment of EPA registration numbers. Assignment of a registration number that differed in active ingredients and/or percent of active ingredient from the true registration number would have resulted in inaccurate conversion of the amount of reported product into pounds of active ingredient. The assignment of generic registration numbers also introduced the possibility of misclassifying a record's product class and/or formulation. Some misassignment of registration numbers is known to have occurred (see section 9.4).
4. Inability to assign EPA registration numbers to all records. Some usage records could not be assigned valid registration numbers based on the information provided by the respondent. These records could therefore not be included in the usage estimates. Similarly, those records assigned a valid registration number which never had a usable amount reported for it also were not included in the estimates. The omission of the amounts reported in those records from the analyses acted to lower the estimates for some active ingredients. This reduction is not believed to have been significant.
5. Errors in state and region level estimates due to proration method. To the extent that the variable under study was not distributed across states served by a reporting aggregate in the same way that the aggregate's total gross sales were, the proration method used on usage data and some other continuous variables contributed to the overall error in the estimates. Although a minority of respondents had to have their data prorated,

there was a marked tendency for respondents who required proration to be large firms, which often reported large amounts of pesticide usage. Therefore, prorated inaccuracies are thought to have had significant impact on the state level usage estimates, and may have affected some region level usage estimates.

6. Respondent measurement error. Usage and other quantitative estimates undoubtedly incorporate some measurement error, but the magnitude of its impact on the estimates is not known.
7. Misinterpretation of questions by respondents. Some specific examples of misinterpretation errors were presented in Chapter 7. Recontact of respondents and logical editing identified and corrected many responses that were originally incorrect due to misinterpretation. Other incorrect responses were either not identified as such, or incapable of correction based on the available information. The impact of this source of error is not known.
8. Inaccurate assumptions underlying nonresponse adjustments. All nonresponse adjustments of the analysis weights assumed that respondents and nonrespondents had similar distributions of characteristics. Distributional differences therefore could result in errors in adjustment. The relatively high coverage rate (see section 9.2) reduced but did not eliminate the potential for significant impact from this source of error.

Once again, it must be stressed that although the various sources of non-sampling error discussed above had the potential to affect the final NUPAS estimates, the degree to which each source acted and the resulting overall effect on the estimates is unquantified and to some extent unquantifiable.

9.4 Quality of the Usage Data

In assessing the final quality of the usage data, a number of factors have to be considered, many of which have been discussed previously (see Chapter 6 and section 9.2). The intent of this section is to provide insight on the degree to which some factors affected the final estimates through the presentation of quantitative descriptions of the usage data. The impact of the process used to assign EPA registration numbers to the usage data is discussed first, followed by an assessment of the impact of subsequent steps in the analysis of the usage data.

In order to assess the accuracy of the EPA registration number assignment process, a simple random sample of 487 records from the final version of the usage file was selected, and those records compared with the corresponding entries on the original questionnaire. RTI personnel who were not involved in the original product data cleaning effort made the comparisons using data sources not available at the time of the original effort. For each record, the following questions were asked.

1. Does the record have a valid EPA registration number associated with it?
2. Was a registration number reported by the respondent?
3. If a registration number was reported, was the record assigned a different number?
4. If a registration number was assigned to the record, is it consistent with the name and formulation information originally provided, and does it seem to contain the same active ingredient(s) in the same percent(s) as the product reported by the respondent?

Of the 487 records in the sample, 20 were never assigned a valid EPA registration number, 122 reported the registration numbers to which they were ultimately assigned, 60 reported legitimate registration numbers but were assigned to different numbers, and the remainder (285) did not report a legitimate registration number and were assigned to one. For the 60 records that reported legitimate numbers but were assigned a different number, comparisons of the original and the assigned numbers lead to the following conclusions:

1. The active ingredients in the two registration numbers were the same in 54 records, and in 45 of these, the percents of those active ingredients were also the same or very similar. Formulation information for the two numbers was similar in 44 of the 54 records.
2. In four records, the two registration numbers contained the same active ingredient, but differed in the presence/absence of another active ingredient. Formulation was similar in all 4 records.

3. The original and assigned registration numbers differed completely in active ingredients in two records.

The records that did not originally report a registration number and were subsequently assigned one can be categorized in the following ways, based on comparison of reported product name and formulation information with the assigned registration number.

1. The assigned registration numbers accurately represented the active ingredients in 268 records, and in 203 of these, the percents of the active ingredients were also accurately represented.
2. In 12 records, the product name reported and that of the assigned registration number were similar, but even in combination with the formulation information reported the name match did not necessarily imply an exact match of active ingredients.
3. Formulation was not reported and the reported product name differed significantly from the product name of the registration number in the remaining 5 records.

Overall, the impression gained from the quality assessment sample was that the registration number assignments accurately represented the reported usage in the great majority of the records. However, in a significant number of records the reported usage was converted into too many or too few pounds of active ingredient; and in a few records it was converted into pounds of an active ingredient not contained in the original reported product.

To appreciate the potential impact of inaccurate assumptions in some of the procedures used to calculate national estimates from the reported usage, the reader should examine the portion of the final estimate that is attributable to each procedure. Table 9-4 shows, for 10 active ingredients, the progression from reported pounds of active ingredient to reported plus imputed pounds to weighted pounds to nonresponse-adjusted weighted pounds.

Table 9-4 Amount of National Usage of 10 Active Ingredients at Four Steps in the Usage Estimation Procedure

Active Ingredient	Number of Obs. with Reported Usage	Number of Obs. with Imputed Usage	Unweighted Pounds Reported	Unweighted Pounds Reported & Imputed	Weighted Pounds ¹	Adjusted Weighted Pounds ²
Chlordane	945	100	1,243,107	1,363,842	1,917,107	3,626,095
Sulfuryl fluoride	55	10	336,824	390,036	1,361,528	2,688,325
Diazinon	2188	193	802,081	824,676	899,105	1,661,691
Methyl bromide	139	22	459,047	518,216	731,588	1,618,251
Dursban	2107	201	534,221	556,904	918,672	1,611,396
Heptachlor	503	41	362,908	376,506	683,034	1,199,746
Atrazine	45	7	111,190	114,718	391,723	1,058,175
Dacthal	80	4	905,992	907,649	892,201	1,027,221
Benlate	239	26	46,799	47,407	29,722	46,182
Zinc Phosphide	168	32	2,030	2,257	3,883	6,965

¹Weighted by the basic weight adjusted for multiplicity.

²Weighted by the basic weight adjusted for multiplicity and nonresponse.

Several conclusions can be drawn from this table.

1. Amount imputation resulted in relatively minor increases in the unweighted number of pounds of active ingredient.
2. The impact of nonresponse adjustment of the weights varied among active ingredients, resulting in increases in the final estimates of between 15 and 121 percent.
3. Multiplicity adjustment had a substantial impact on some of the final estimates. This is especially noticable for dacthal and benlate, for which the multiplicity factors in some records were large enough to more than offset the basic weights, resulting in weighted estimates that were smaller than the original, unweighted reported amounts. In most cases, the largest multiplicities were associated with multi-location, multistate companies which reported large amounts of usage. Therefore, any inaccuracies in multiplicity estimation due to errors in mapping the reporting aggregate to the frame in mail states and/or ascertaining the number of employees in telephone states who were on the regulatory lists in those states had the potential for significant impact on the final usage estimate. Overestimation of telephone state multiplicity was a likely cause for the reduction of dacthal and benlate amounts.

9.5 Quality of the Nonusage Data

Assessment of the quality of the nonusage data collected by NUPAS must be conditioned on the intended relationship between the nonusage data and the usage data. The major goal of NUPAS was to collect and analyze pesticide usage data, and nonusage data was sought primarily to assist in and enhance analyses of the usage data. Questionnaire design was governed by this, as were the reporting aggregate resolution efforts.

The quality of the nonusage data varies considerably among questions. Some questions were particularly prone to respondent misinterpretation (see section 7.2), and some, most notably the questions on certified/licensed employees, have definitional problems which make interpretation of responses difficult. Other questions (e.g., trade association membership and trade magazine subscriptions) were relatively straightforward both to answer and interpret. Levels of nonresponse varied among questions. The

highest levels were for questions perceived to be sensitive (e.g., sales information) or difficult to interpret (e.g., number of employees). Efforts were made during the reporting aggregate resolution callbacks to obtain response to unanswered questions that were particularly important for the usage analyses (e.g., questions on total gross sales (question 18) and industry sectors served (question 21)).

The fact that some firms provided aggregate data affected both data quality and interpretation. Data quality was adversely affected when a respondent answered different questions at different levels of aggregation, and some questions (e.g., how many employees are certified/licensed in the reference state?) could not be appropriately answered by some respondents because of the level of aggregation they chose to respond at.

9.6 Overall Strengths and Weaknesses of the NUPAS Data

Most of the discussion in the preceding sections of this chapter has dwelled on the potential/realized errors incorporated in the NUPAS database. Despite these problems, the NUPAS database has sufficiently many strengths to make it a useful resource. This survey remains the most comprehensive national survey of pesticide use by PCO's in the tree, lawn, and structural sectors done to date. A large fraction of the firms in these sectors were sampled, and firms of all sizes and in all areas of the country are represented in the database. The use of a probability sample design permits the calculation of variance estimates.

Those who use the NUPAS database and tabulations should be especially aware of the following considerations when interpreting the estimates.

1. The data is for 1981. Since then, some industry sectors have grown, and products may have both come into and gone out of use.
2. Estimates for categorical variables are in terms of number of reporting aggregates. There is no standard definition for reporting aggregate.

3. National estimates are less prone to be biased than regional or state estimates, since sample sizes are larger and proration by percent of sales is not necessary.

Overall, the NUPAS database is the best source available for national and regional information on pesticide use by the tree, lawn, and structural pest control sectors. It is based on sound statistical and survey methodology and a large sample. Considerable effort was expended to assure that the data quality is as high as possible, given that this survey was the first of its kind. The database should prove useful to a wide variety of pesticide users and regulators, and serves as an excellent base for future surveys of the pest control industry.

10. THE ANALYSIS PLAN FOR NUPAS DATA

Decisions on the types of analyses to be included in the NUPAS final report were based on the following five considerations:

1. The quality of the data
2. The needs of EPA
3. The needs of the sponsoring trade associations
4. The content of the user data file
5. Cost and schedule constraints.

In general, the goal of the analysis plan is to present results of interest to the survey sponsors, to pesticide users, and regulators. Decisions on which variables to include in the analyses, and the levels at which to represent them, were influenced by known limitations of the data. Care was taken not to overextend the data. Site- and sector-specific analyses of the usage data were not included in the analysis plan because they can be done by EPA or the industry sponsors using the information on the user data file.

The final analysis plan, agreed to by RTI, EPA, and the industry sponsors in February 1984, is presented in Appendix D. Some minor changes have been made in the plan since that time, and unanswered questions have been resolved. A decision was made to partition the original tree/lawn sector category into tree only, lawn only, and tree and/or lawn subsectors, based on questionnaire data. Respondents in the tree/ lawn sector were assigned to the subsectors in the following way.

1. If usable data was provided in question 21, respondents which reported a ratio of (% tree, shrub, ornamental)/(% lawn, turf) of greater than 10 were assigned to the tree only subsector. If the ratio of (%lawn, turf)/(% tree, shrub, ornamental) was greater than 10 then they were assigned to the lawn only subsector. If neither ratio exceeded 10, they were assigned to the tree and/or lawn subsector.

2. Respondents which did not have usable question 21 data, but did answer the question on national trade association membership, were assigned to the tree only subsector if they reported belonging only to the NAA, to the lawn only subsector if they reported belonging only to the PLCAA, and to the tree and/or lawn subsector otherwise.
3. Respondents which did not provide usable answers to either of the above questions were assigned to the tree and/or lawn subsector.

APPENDIX A
Edit Manual for NUPAS

National Urban Pesticide Applicator Survey
EDIT PROCEDURAL MANUAL

A. Introduction

The National Urban Pesticide Applicator Survey (NUPAS) is sponsored by the Environmental Protection Agency (EPA) and the national organizations that are associated with general pest control maintenance. The survey is being conducted by mail with telephone follow-up (if necessary) of selected nonrespondents.

Your job as an editor is extremely important to the overall success of this project. You must review each questionnaire and make editorial decisions based on the information given to you at your training and in this manual. Use this manual throughout your work on the project; make any notations in it that may be helpful.

B. Your Supervisor

Your direct technical supervisor for this operation is Evelyn Scalf who will be assisted by ~~Dick Waddell~~. See either Evelyn or ~~Dick~~ if you have questions or problems.

C. General Edit Guidelines

This project does not differ greatly from others; therefore, you should be familiar with most of the general edit guidelines. They are listed below:

- Use only blue pencils when making editorial notations in the questionnaires.
- The following standard codes are being used:

MR = Multiple Response
BD = Bad Data (Illegible, etc.)
RE = Refusal
DK = Don't Know (unless specific response category furnished for this response)
BL = Blank.

- FAIL EDIT cases must be accompanied by an Edit Problem Sheet; these cases go directly to your supervisor.
- We are not coding "Other (SPECIFY)" responses. However, try to "force up" these responses to a precoded response if possible.
- If respondent circled actual answer rather than code, circle corresponding code.

- Use red pen when transferring to original questionnaires all info collected ^{by AARP.} A-I-P

D. QxQ Edit Specs

General Specifications

Error Resolution

Section A:

- | | |
|--|---|
| <p>1. Only one code should be circled.</p> <hr/> | <ul style="list-style-type: none">• If marginal notes, see supervisor. <hr/> |
| <p>2. Should be completed if 02 was circled in Q. 1.</p> | <ul style="list-style-type: none">• If changes were made on label and not entered in box, transfer information to box.• Clarify entries for legibility.• If 01 is circled in Q. 1 and Q. 2 is answered, see supervisor.• Standard two alpha character state code should be entered. (See List A.) <hr/> |
| <p>* 3. Only one code should be circled.</p> | <ul style="list-style-type: none">• If <u>original</u> response is 02, ~~check rest~~ check rest of questionnaire for additional entries before coding.• Responses must be recoded as follows and new codes entered in the margin to the right of arrow line.<ul style="list-style-type: none">01 = Yes02 = No; no other entries in questionnaire03 = No; additional entries in questionnaire• Line through original response codes. <hr/> |
| <p>4. Length of time in years and/or months should be given.</p> | <ul style="list-style-type: none">• OVO Box -- Enter number of years in the <u>first 3 boxes</u> and the number of months in the <u>last two boxes</u>.• If fractions are given, convert to years and/or months (1½ years, enter 001 year and 03 months). <hr/> |
| <p>5. One number should be entered.</p> | <ul style="list-style-type: none">• If a marginal note that changes the response is entered, FAIL EDIT. <hr/> |

General Specifications

6. Only one code should be circled. If code 01 or code 03 is circled, a number should be entered after arrow to right of codes.

7. Only one code should be circled.

8. Only one code should be circled.

9. Only one code should be circled.

Error Resolution

- If a number is entered for "How many" and the code is not circled, circle corresponding code.
- If 02 and 03 are circled, use 02 and line through 03.
- If there are other multiple responses, code MR in the right margin and circle.

- If code 03 is circled and a license number is shown on the label, line through response and code BD.

Section B:

10. All answers must be whole numbers.

Do not put "administrative" in d.

Self - 10.5 - 1.0

- If whole numbers are given, add .0 (e.g., if answer is 2, edited entry should be 2.0).

If fractions, ranges, or decimals are given, FAIL EDIT.

If "f" is specified as one of the previous categories, FAIL EDIT.

- If there are marginal comments but no numeric entries, FAIL EDIT.

11. Answer must be whole number.

Self - 1.0

- If whole numbers are given, add .0 (e.g., if answer is 2, edited entry should be 2.0).

If fractions, ranges, or decimals are given, FAIL EDIT.

12. One code should be circled.

General Specifications

Error Resolution

- | | |
|--|---|
| 13. One or more codes should be circled. | • If 03 is circled and respondent enters the definition of reciprocity (<u>reciprocity</u> : the recognition by one or more states of the validity of a license granted by another state) line through 03 and code 02. |
| <hr/> | |
| 14. One or more codes may be circled. | • If state organization is listed under "other", see supervisor. |
| <hr/> | |
| 15. Only one code should be circled. | |
| <hr/> | |
| 16. No edit required. | |
| <hr/> | |
| 17. One or more codes should be circled. | |
| <hr/> | |

Section C:

- | | |
|---|--|
| C. Circle box on last line of instructions for the benefit of Data Entry. | • If box marked, enter "A" below box and circle; if box not marked, enter "B" below box and circle. |
| <hr/> | |
| 18. One code should be circled. | • OUO Box -- If marginal notes are given for <u>questions 18 through 21</u> , code <u>AN</u> . If no notes given, code <u>BL</u> . |
| <hr/> | |
| 19. One code should be circled. | |
| <hr/> | |
| 20. Categories should be answered with whole numbers. | • If fractions or decimals are given in either column, round up.

• If ranges are given, take mid-point. |
| <hr/> | |
| 21. Categories should be answered with whole numbers. | • If fractions or decimals are given in either column, round up.

• If ranges are given, take mid-point. |
| <hr/> | |

General Specifications

Error Resolution

Section D:

22. Circle box in last line of instructions for the benefit of Data Entry.

- If box marked, enter "A" below box in circle; if box not marked, enter "B" below box in circle.

- If respondent refuses, enter R ^{under} in box: *circle*

- If entries were made and marked through, renumber good entries so that no number is missed.

- Enter the total number of lines used in listing products in the left margin beside table headings.

Col. 1 (Product Name)

- If illegible, code BD.

Col. 2 (Formulation)

Col. 3 (EPA Reg. No.)

Col. 4 (Ready to use or Concentrate)

- Only one code should be circled. If both "R" and "C" are circled, code MR.

Col. 5 (Total Amount)

- If entry in Col. "Total Amt." is 100,000,000 or more, FAIL EDIT.
- If fractions or decimals are given in col. 5, round up.
- If ranges are given, take mid-point.
- Line through gallons, pounds, etc. and enter the appropriate code (see below) in the right margin.

Ounces	= A	Quarts	= G
Pounds	= B	Gallons	= H
Grams	= C	Liters	= J
Kilograms	= D	Insects	= K
Fluid Ounces	= E	Other	= L
Pints	= F		

- If measurements other than those listed above are given, FAIL EDIT.

General Specifications

Error Resolution

22. (Continued)

- If no more than 28 products are listed and the box at the bottom of p. 8 is not checked, code "B" above box and circle.
- If the box at the bottom of page 8 is checked and additional sheets are included, transfer information to supplemental sheets and staple to the top of p. 9.
- Code "A" in box on last sheet, circle, and line through box on p. 8.
- Be sure to enter ID # and number products on supplemental pages.

Section E:

23. One code should be circled.

24. Dates should be given if 02 was circled in Q. 23.

- Recode months and years numerically in OVO boxes.

25. One or more codes should be circled.

- If entry is made under "other" and nothing is circled, circle 03.

26. May not be answered.

- If response entered, code A in box; if no response, code B.

* 27. Only one code should be circled.

make a list of
IPth or L. B. L. #
of every individual in

28. Name and phone number should
entered.

- Enter the following codes in the right margin, as applicable:

if black = 13
Name only = NO

Telephone only = T0

Both = BO

A-6-

APPENDIX B

Reporting Aggregate Resolution Questionnaire

ID

--	--	--	--	--	--	--

NATIONAL URBAN PESTICIDE APPLICATORS SURVEY
Weighting Resolution Questionnaire

INTRODUCTION

Hello, my name is _____ and I am calling from the Research Triangle Institute in North Carolina. You were identified as the contact person on a questionnaire returned by your firm for the National Urban Pesticide Applicators Survey.

We are beginning to analyze the data from this survey and would like to clarify some of the information in the questionnaire.

SECTION A

(ASK THIS SECTION WHEN Q5 > 1)

- A1. It was reported that (FIRM NAME) has (NUMBER IN Q5) business offices or locations in (REF. STATE). Is that correct?

Yes	01	(A2)
No.	02	
DK.	94	] (A2)
RE.	96	

- a. What is the correct number?

--	--

- A2. Does your firm operate any other offices or locations outside the state of (REF. STATE) under the name (FIRM NAME)?

Yes	01	
No.	02	
DK.	94	] (A3)
RE.	96	

- a. In what states are the other offices located? How many separate offices are in (STATE)?

STATE	NUMBER				
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>			<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>		
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>			<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>		
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>			<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>		
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>			<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>		
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>			<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>		
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>			<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td> </td><td> </td></tr></table>		

ID

--	--	--	--	--	--	--

A3. Does the pesticide product usage information provided in the questionnaire pertain to all (NUMBER IN A1 + A2) locations or just to (ADDRESS) or to some other grouping of the locations?

All locations 01
Address of quex location. . . 02
Other (SPECIFY) 03

DK. 94
RE. 96

ID

--	--	--	--	--	--	--

SECTION B

(ASK THIS SECTION WHEN Q6 = 01 OR DK OR MISSING, OR IF A2 = YES.)

B1. IF Q6 IS MISSING OR DON'T KNOW, OR IF A2 = YES, THEN ASK THE FOLLOWING, OTHERWISE GO TO B2. Does your firm provide pest control services to a single state area or to a multi-state area?

Single state.	01	(SECTION C)
Multi-state area.	02	(B3)
DK.	94	
RE.	96	(SECTION C)

B2. It was reported that (FIRM NAME) provided service in (NUMBER IN Q6) states. Is that correct?

Yes	01	(B3)
No.	02	
DK.	94	
RE.	96	

a. What is the correct number of states?

--	--

DK.	94
RE.	96

B3. In what states is service provided?

a. Approximately what percent of the firm's total business is in each state?

STATE		%		STATE		%																				
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>					. . .	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>							. . .	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>					. . .	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>						
<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>					. . .	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>							. . .	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>					. . .	<table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table>						

b. What is this percentage based on?

Gross revenues or sales . . .	01
Total number of accounts. . .	02
Other (SPECIFY)	03

DK.	94
RE.	96

ID

--	--	--	--	--	--	--	--

B4. Does the pesticide product usage information provided in the questionnaire pertain to all states served or just for services in (REF. STATE)?

Reference state only. 01
 All states. 02
 Other (SPECIFY) 03

DK. 94
 RE. 96

B5. FOR EACH STATE SERVED (AS IDENTIFIED IN B3) THAT IS IN THE LIST BELOW ASK THE FOLLOWING. OTHERWISE GO TO NEXT APPROPRIATE SECTION. How many persons in your firm are certified to apply pesticides in (STATE)?

<u>STATE</u>	<u>NUMBER</u>	(IF NONE IN A STATE ENTER 00.)
Arizona	_____	
Louisiana	_____	
Massachusetts	_____	
Nebraska.	_____	
Pennsylvania.	_____	
Virginia.	_____	
West Virginia	_____	

ID

SECTION C

(ASK THIS SECTION WHEN Q7 = 01.)

- C1. Does the ownership of (FIRM NAME) offer pest control services under any other business name?

Yes 01 (C3)
No. 02
DK. 94
RE. 96

- C2. IF SECTION A WAS NOT USED THEN ASK THE FOLLOWING. OTHERWISE GO TO C4.
Does the ownership of this firm offer pest control services through any other offices or locations?

Yes 01
No. 02
DK. 94 (GO TO SECTION D)
RE. 96

- a. How many business offices or locations does your firm have?

(C4)

- C3. What is the name and address of each such firm?

NAME	ADDRESS
_____	_____
_____	(Street)
	(City) (State) (ZIP)
_____	(Street)
	(City) (State) (ZIP)

- C4. Does the pesticide product and usage information you provided in the questionnaire pertain to only (FIRM NAME) at (ADDRESS) or to something else?

Named firm only 01
Other (SPECIFY) 02

DK. 94
RE. 96

ID

--	--	--	--	--	--	--

C5. IF SECTION B HAS NOT BEEN COMPLETED AND Q6 IS MISSING OR DK THEN GO TO SECTION B. OTHERWISE ASK: The questionnaire reported that your firm provides pest control services only in (REF. STATE). Is that correct?

Yes	01	(SECTION D)
No.	02	
DK.	94	
RE.	96	(SECTION D)

a. How many states does your firm serve?

--	--

 (B3)

ID

--	--	--	--	--	--	--

SECTION D

(ASK ALL APPROPRIATE QUESTIONS. REMIND THE RESPONDENT TO ANSWER THESE QUESTIONS SO THEY CORRESPOND TO THE UNIT FOR WHICH PESTICIDE PRODUCT USAGE INFORMATION WAS GIVEN.)

- D1. IF SECTION B WAS NOT COMPLETED AND THE REFERENCE STATE IS ANY ONE OF ARIZONA, LOUISIANA, MASSACHUSETTS, NEBRASKA, PENNSYLVANIA, VIRGINIA, OR WEST VIRGINIA THEN ASK THE FOLLOWING. OTHERWISE GO TO D2.

[The questionnaire returned to us reported that (NUMBER IN Q11) persons are certified or licensed to apply pesticides in (REF. STATE).] How many persons in (your firm/the firms you are reporting for) are certified to apply pesticides in (REF. STATE)?

STATE

--	--

 NUMBER

--	--

- D2. IF Q21 WAS NOT ANSWERED ASK THE FOLLOWING, OTHERWISE GO TO D3. Approximately what percent of the gross sales for a year comes from (CATEGORY BELOW)?

Lawn and turf	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>				%
Tree, shrub and ornamental.	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>				%
Termites and other wood destroyers.	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>				%
Commodity fumigations	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>				%
General pest control.	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>				%
Other (e.g., pesticide sales)	<table border="1"><tr><td> </td><td> </td><td> </td></tr></table>				%

- D3. IF Q18 WAS NOT ANSWERED ASK THE FOLLOWING, OTHERWISE GO TO D4. Which of the following categories includes the total gross sales figure for all activities?

Less than \$25,000	01
\$25,000 - \$49,999	02
\$50,000 - \$99,999	03
\$100,000 - \$199,999	04
\$200,000 - \$499,999	05
\$500,000 - \$999,999	06
\$1,000,000 - \$5,000,000	07
Over \$5,000,000	08
DK.	94
RE.	96

ID

--	--	--	--	--	--	--

D4. IF SECTION A IDENTIFIES MORE THAN ONE LOCATION FOR THIS FIRM, ASK THE FOLLOWING. OTHERWISE GO TO END. What (is/are) the address(es) of the (NUMBER IN A1 + A2) location(s)? (DO NOT INCLUDE THE LOCATION GIVEN ON THE QUEX.)

	STREET	CITY	STATE	ZIP
(1)	_____	_____	_____	_____
(2)	_____	_____	_____	_____
(3)	_____	_____	_____	_____
(4)	_____	_____	_____	_____
(5)	_____	_____	_____	_____

END

I have no other questions at this time. Thank you very much for taking the time to clarify this information for our survey.

ID

--	--	--	--	--	--	--

SECTION E

(USE THIS SECTION ONLY WHEN Q8 = 1, 2, OR 3.)

E1. It was reported that sometime in the past 18 months your firm (merged with/split from/merged with and split from) another firm or firms. Is that correct?

Yes, merged	01	(E3)
Yes, split.	02	(E4)
Yes, both	03	(E3)
No.	04	
DK.	94	
RE.	96	(D4)

E2. Did your firm (merge with/split from) any other firms sometime in the past 18 months?

Yes, merged	01	
Yes, split.	02	(E4)
Yes, both	03	
No.	04	
DK.	94	
RE.	96	(D4)

E3. What are the names and addresses of the firms that you merged with?

NAME	ADDRESS
	(Street)
	(City) (State) (ZIP)
	(Street)
	(City) (State) (ZIP)

(IF THE FIRM DID NOT SPLIT, THEN GO TO E5.)

ID

--	--	--	--	--	--	--

E4. What are the names and addresses of the firms that were involved in the split?

NAME	ADDRESS
	(Street)
	(City) (State) (ZIP)
	(Street)
	(City) (State) (ZIP)

E5. Does the pesticide product usage information provided in the questionnaire pertain to all firms involved in the (merger/split/merger and split) or only to (FIRM NAME) at (ADDRESS) or to something else?

- All firms 01 (E6)
 Firm on quex only 02 (D4)
 Other (Specify) 03 (E6)

E6. IF THE PESTICIDE PRODUCT USAGE DATA IS REPORTED FOR MORE THAN ONE STATE AND SECTION B HAS NOT BEEN COMPLETED THEN ASK THE FOLLOWING. OTHERWISE GO TO D4. In what states do these firms provide pest control services?

a. Approximately what percent of these firms' total business is in each state?

STATE		%		STATE		%
	. . .				. . .	
	. . .				. . .	
	. . .				. . .	

(GO TO D4)

APPENDIX C

Letter to Dr. Richard Carr Outlining Information Required
for Interpretation of Certification/Licensing Questions
and Definition of Realized Target Population

RESEARCH TRIANGLE INSTITUTE

Center for Survey Statistics

February 15, 1984

Dr. Richard V. Carr
P. O. Box 3313
McLean, VA 22103

Dear Dick:

As we indicated at the February 6 meeting with EPA, our interpretation of answers to the questions asked about certification in NUPAS is hampered by a lack of detailed understanding of the nature of the licensing/certification programs in 1981 in each of the states covered by NUPAS. Such an understanding is also needed to allow us to fully define the target population addressed by the survey. The intent of this letter is to specify the sort of information that we require for resolution of our questions, and to ask for your assessment of the feasibility of obtaining such information given the schedule and budget constraints we are working under.

For each state, we need information on the following:

- certification of individuals who apply pesticides
 - is it necessary for all individuals or only owners or supervisory personnel?
 - is it necessary for all pesticides or only restricted use pesticides?
 - to what degree is participation voluntary (i.e., do people who do not legally have to be certified obtain certification anyway for professional reasons)?
 - how often does certification have to be renewed?
 - how frequently are lists of certified people updated (i.e. expired certifications removed, new ones added, addresses and telephone numbers updated)?
- licensing of individuals who apply pesticides
 - is licensing different from certification? if so,

- is it necessary for all such individuals, or only for a subset of them?
- is it necessary for all pesticides or only for restricted use pesticides?
- to what degree is participation voluntary?
- how often do licenses have to be renewed?
- how frequently are lists of licensed people updated?
- licensing of firms that offer commercial pest control services
 - is there any? if so,
 - is it done for each location of the firm operating within the state, or only for a main office of the firm?
 - is it mandatory for any pesticide use, or only for restricted pesticide use?
 - can/do some firms choose to become licensed although not legally obligated to?
 - how often do licenses have to be renewed?
 - how frequently are lists of licensed firms updated?
- reciprocity of regulation with other states
 - is there any? if so,
 - with what other states?
 - at what level (i.e., individual certification, individual licensing, firm licensing)?
 - are reciprocity regulatees included on state lists?
- industry classification on the list(s) provided by the state
 - how well does it coincide with federal categories 3a, 3b, and 7?
 - how are specializations (e.g. commodity fumigation, right-of-way pest control) classified by the state?
 - did the lists as provided include or exclude any 3a, 3b, or 7 applicators/firms because of the state's classification system?

Dr. Richard V. Carr
February 15, 1984
Page 3

• changes in the regulatory programs since 1981.

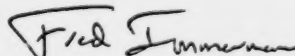
At present, our knowledge of the state regulatory programs is limited to the information contained in the two attached tables.

Ideally, we would like to have a report that describes the regulatory program in each state, and evaluates how well the sampling frame we developed from the state list covers the population of interest to this survey. Such a report would be based primarily on documentation obtained directly from each state, enhanced by (or perhaps replaced by, if necessary) your evaluations and interpretations, with deviations from strict documentation clearly noted.

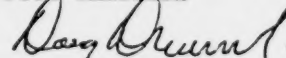
As you know, RTI is committed to an April 15 completion date for the NUPAS final report, and so the information outlined above will have to be gathered within a relatively short period of time if it's to be of use. Dollar resources are also finite, and cannot support an extended data gathering effort. One option for reducing the impact of these constraints is to provide research and secretarial support for you through RTI's Washington office.

We would appreciate acknowledgement of receipt of this letter, at which time we could attempt to answer any initial questions you might have, and agree upon a date by which you could provide us with your assessment of the feasibility and cost (in terms of time and dollars) of producing the report described above. We currently estimate that this assessment of feasibility would require one day of your time, for which you would be paid at your (soon to be approved) EPA rate. If the project seems to have a reasonable chance of enhancing the use of the NUPAS data and if costs are not projected to be prohibitive, we would then authorize you to proceed. We look forward to hearing from you.

Sincerely,



Fred Immerman



Doug Drummond

FI/pp

cc: Peter Kuch
Edward Brandt

Table 1. Licensing vs. Certification in the Seven "Telephone" States

<u>State</u>	<u>Licensing/Certification Structure</u>
Arizona	State licenses are issued to those companies with "services for hire." Owners and applicators within companies or services must be certified by examination to use restricted use pesticides.
Louisiana	Same As Arizona.
Massachusetts	To use restricted pesticides on your own or as a company or service, an exam is required for certification. An individual may obtain an applicator's license to use restricted pesticides under the supervision of a certified applicator.
Nebraska	Applicators of restricted use pesticides must be certified. There is no licensing.
Pennsylvania	Certification and licensing are synonymous in this state. An exam is required.
Virginia	A test must be passed to become certified, with recertification every 3 years. Upon certification, a license is obtained annually for a fee. A person may be licensed without certification only for nonrestricted pesticides.
West Virginia	Certification requires passing an examination. Recertification is required every three years after completing a training session. Licenses are obtained annually with a guarantee of certification, insurance, and fee. Companies or services with certified applicators are required to obtain licenses. Certified applicators may work under the company license number.

Note: The above information was gathered by a telephone by a person who was not knowledgeable about pesticides or regulatory programs. As such, the information would (ideally) be verified and refined.

Table 2. Nature of Licensing Lists by States

State Name	Licensing Category-1/	Unit Licensed Firm-2/ Individual-3/	Correspondence of Licensed Category-4/	Form of List-5/		Entries on Frame		Source of List
				Original	Cleaned	Original	Cleaned	
Alabama	3 7	6/6/	2 2	2 2		880	532	St. Dept. of Agric.-Plant Div.
Arizona	3 7	2/3/	1 1	2 2		1200	~2/	St. Dept.-Pest. Control Bd.
Arkansas	3 7	✓	2 2	2 2		520	266	St. Dept. of Commerce-Plant Div.
California	3 7	2/2/	3 3	1,2 1,2		2100	2170	St. Dept. of Food & Agric.
Colorado	3 7	✓	2 2	2 2		312	264	State Dept. of Agric.
Connecticut	3 7	✓	2 2	1 1		1100	836	State Dept. of Envir. Protection
Delaware	3 7	✓	2 2	2 2		168	132	State Dept. of Agric.-Production & Promotion
District of Columbia	3 7	6/6/	1 1	2 2		195	142	Council of D.C.
Florida	3 7	2/2/	3 3	2 2		1298	1254	Dept. of Health & Rehabilitative Services
Georgia	3 7	6/6/	1 1	1,2 1,2		1000	669	State Dept. of Agric.
Hawaii	3 7	✓	1 1	1,2 1,2		250	191	State Dept. of Agric.

Reproduced from
best available copy.

Table 16. Nature of Licensing Lists by States (cont.)

State Name	Licensing Category 1/	Unit Licensed Firm 2/	Individual 3/	Correspondence of Licensed Category 4/	Form of List 5/		Entries on Frame		Source of List
					Original	Cleaned	Original	Cleaned	
Idaho	3 7		✓ ✓	2 2	3 3		406 1300	91 1167	State Dept. of Agric.-Plant Div. Dept. of Public Health
Illinois	3 7	6/ 6/		1 1	1 1				
Indiana	3 7	6/ 6/		1 1	4 4		706	695	State Chemical & Seed Commission
Iowa	3 7	6/ 6/		1 1	1 1		1170	371	State Dept. of Agric.-Pesticide Div.
Kansas	3 7	6/ 6/		1 1	3 3		825	449	State Bd. of Agric.-Pesticide Div.
Kentucky	3 7		✓ ✓	1 1	2 2		440	405	Sent thru Grimes' office/no cover letter
Louisiana	3 7	2/ 2/	✓ ✓	3 3	1 2		780	-3/	State Dept. of Agric.
Maine	3 7	2/ 2/		3 3	2 2		115	85	State Dept. of Agric. Food & Rural Sources
Maryland	3 7	6/ 6/		2 2	3 3		576	459	State Dept. of Agric.
Michigan	3 7	2/ 2/		3 3	2,3 2,1		1776	845	Dept. of Agric.-Plant Industry
Minnesota	3 7		✓ ✓	1 1	1 1		2900	239	State Dept. of Agric.-Dept. of Agronomy

Table 16. Nature of Licensing Lists by States (cont.)

State Name	Licensing Category-1/	Unit Licensed Firm-2/ Individual-3/	Correspondence of Licensed Category-4/	Form of List-5/		Entries on Frame		Source of List
						Original	Cleaned	
Mississippi	3 7	✓ ✓	2 2	2 2		316	309	Dept. of Agric.-Plant Div.
Massachusetts	3 7	✓ ^{3/} ✓ ^{3/}	3 3	1 1		4900	-3/	Dept. of Agric.
Missouri	3 7	✓ ✓	1 1	1 1		1296	1058	State Dept. of Agric.-Pesticide Control
Montana	3 7	✓ ✓	2 2	1 1		215	100	Dept. of Agric.-Environ. Mngt. Div.
Nebraska	3 7	✓ ^{3/} ✓ ^{3/}	3 3	1 1		644	-3/	Air and Waste Compliance Bd.
Nevada	3 7	✓ ^{2/} ✓ ^{2/}	3 3	2 2		108	78	State Dept. of Agric.
New Hampshire	3 7	✓ ^{6/} ✓ ^{6/}	2 2	2 2		220	94	State Dept. of Agric.-Pest Control Div.
New Mexico	3 7	✓ ✓	1 1	1 1		200	170	State Dept. of Agric.
New York	3 7	✓ ^{6/} ✓ ^{6/}	2 2	1 1		5561	1795	N.Y. Dept. of Environmental Conserv.
New Jersey	3 7	✓ ^{6/} ✓ ^{6/}	2 2	1 1		1417	1218	Dept. of EPA-N.J.
North Carolina	3 7	✓ ✓	3 3	1 1		3372	720	Dept. of Agric.-Pesticide Div.

Table 16. Nature of Licensing Lists by States (cont.)

State Name	Licensed Category 1/	Unit Licensed Firm 2/	Individual 3/	Correspondence of Licensed Category 4/	Form of List 5/	Entries on Frame Original	Cleaned	Source of List
Ohio	3 7		✓ ✓	3 3	3 3	3625	1251	State Dept. of Agric.
Oklahoma	3 7	6/ 6/		1 1	2 2	1350	727	State Dept. of Agric.-Plant Div.
Oregon	3 7	6/ 6/		1 1	1 1	477	265	State Dept. of Agric.
Pennsylvania	3 7		3/ 3/	3 3	1 1	2500	3/	State Dept. of Agric.
Rhode Island	3 7		✓ ✓	1 1	2 2	378	124	Div. Environmental Management
South Carolina	3 7		✓ ✓	2 2	1 1	832	571	State Dept. of Agric.
South Dakota	3 7		✓ ✓	2 2	2 2	630	84	State Dept. of Agric.
Tennessee	3 7	6/ 6/		1 1	3 3	484	336	Dept. of Agric.-Plant Div.
Texas	3 7	6/ 6/		2 2	1 1	3600	2122	Dept. of Agric.
Utah	3 7		✓ ✓	2 2	2 2	104	98	Dept. of Agric.
Vermont	3 7		✓ ✓	1 1	2 2	168	72	Dept. of Agric.

Table 16. Nature of Licensing Lists by States (cont.)

State Name	Licensed Category 1/	Unit Licensed Firm 2/	Individual 3/	Correspondence of Licensed Category 4/	Form of List 5/	Entries on Frame Original	Cleaned	Source of List
Virginia	3 7	√ √	√ √	3 3	1 1	1507	-3/	Dept. of Agric.-Consumer Div.
Washington	3 7	√ √		2 2	2 2	196	180	Dept. of Agric.
West Virginia	3 7	√ √	√ √	3 3	4 4	137	-3/	Dept. of Agric.
Wisconsin	3 7	√ √		2 2	1 1	375	220	Dept. of Agric.-Trade & Consumer Div.
Wyoming	3 7	√ √		3 3	2 2	50	47	Dept. of Agric.

1/Tree/Lawn = Cat. 3
Structural = Cat. 7

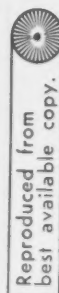
2/Indicates license numbers had to be obtained by field staff.

3/Indicates firm name not given. A sample of individuals were contacted by telephone.

4/Category Codes 1=same 2=some amount of interpretation 3=large amount of interpretation.

5/Form of List 1-Printout 2-Letter 3=Book 4=Hailing labels.

6/Firm license number was available for entire frame and was used to identify unique firms.



APPENDIX D

Memo to EPA on the NUPAS
Analysis Plan and User File Content

RESEARCH TRIANGLE INSTITUTE

Center for Survey Statistics

February 21, 1984

MEMORANDUM

TO: Edward Brandt

FROM: Fred Immerman *FI*
Doug Drummond *DD*

SUBJECT: Final Analysis Plan for NUPAS

This memo describes our understanding of the final NUPAS analysis plan agreed to in our meeting with you, Linda Zarow, Peter Kuch, Jack Grimes, and Dick Carr in D.C. on February 6, 1984 and subsequently discussed in our telephone conversation of February 16, 1984.

A. General Points

A draft version of the final report will be prepared for an April 2 meeting with EPA and trade association representatives. Sections prepared by EPA and the trade associations should be ready at this time. These sections should cover the following topics:

- why the group chose to be included in NUPAS
- what the group expected from NUPAS
- perception of what came out of NUPAS (i.e. what questions were answered, what questions still remain, what the quality of the NUPAS data is)
- how NUPAS can be used, including effect of changes in industry since 1981.

The final report and a user file will be delivered to EPA on April 16, 1984. The analysis plan assumes that the user file will contain reporting aggregate by state level data from questions 14, 20, 21, and 22, as well as geographic data that will permit state-level estimates of usage to be calculated. Therefore, detailed cross tabulations of questions 14, 20, 21, and 22 will not be included in the final report, but will instead be left to EPA to generate from the user file. Only data from eligible respondents will be analyzed. Determination of eligibility will be discussed in detail in Volume 2 of the documentation. Analyses in the final report will be at the national or EPA region level. If any state level estimates are gene-

rated from the NUPAS data, they will be presented in separate reports. Variance estimates will be generated only for analyses involving usage data. It will not be possible for analysts to correctly generate variance estimates from the user file.

B. Usage Data

In the final report, estimated total quantities of the top 10 active ingredients used in each pesticide class and overall will be profiled by formulation of product, industry category, and EPA region as shown in Tables 1-3. A table of total pounds of a.i. used by industry category and region will also be produced for each pesticide class. Listings by class of estimated total use for all active ingredients reported in NUPAS will be given in an appendix to the final report. Pesticide class here refers to broad categories of pesticides (i.e. insecticides (including nematocides), fungicides, herbicides, microbicides, and rodenticides (including repellents and poisons for other warm-blooded non-human vertebrates)). Pesticide class will be designated at the product level based on the type codes and/or product manager number on the EPA product label file. Some active ingredients will be represented in more than one pesticide class. Formulation classification will also be done using the EPA product label file. Because there are 20 distinct EPA formulation codes on this file (Table 4), some of which are not pertinent or are too specific for our purposes, there should be some effort to reduce these codes to fewer, more general formulation categories. We suggest the following seven categories:

1. granular or pelleted
2. dust
3. wettable powder
4. concentrate
5. ready to use solution
6. pressurized
7. other

The distribution of original formulation codes for the 1073 products reported on in NUPAS is given in Table 5a, and the distribution for the generalized codes is shown in Table 5b.

All sample members were originally classified into one of four industry categories (tree/lawn, structural, both, or unknown) based on information on the state lists used to construct the NUPAS sampling frames. We subsequently reclassified respondents based on the answers provided in questions 21, or, in lieu of data in question 21, question 14. Those respondents who did not provide usable data in questions 14 or 21 were left in their original industry category unless that category was "unknown". Hard copy was checked for such respondents, and they were assigned to one of the informative categories based on their product data and/or answers to question 17 and/or firm name. Industry type is always designated on an overall basis (i.e. not on a region or state level basis for those aggregates serving more than one state).

MEMORANDUM

Page 3

February 21, 1984

It is possible to create a more specific industry classification for respondents, using the same data discussed above. Such a classification could have the following categories: tree, lawn, tree and/or lawn, structural, structural and tree, structural and lawn, and structural and tree and/or lawn. As we discussed, this more specific classification may be preferable for some or all of the analyses involving industry categories. Further decisions on this will await generation of some preliminary tables with raw cell counts.

Each cell in Tables 1-3 will contain the following information:

- the estimated total amount of active ingredient, in pounds
- the estimated standard error of the total, and
- the weighted number of reporting aggregates that reported use of this active ingredient.

Both the name and the EPA chemical code of the active ingredient will be given.

C. Other Data

Economic profile data (questions 18 and 19) will be presented for the nation and for each EPA region by industry category (see Tables 6 and 7). Sales categories will be based on the collapsed intervals used in the interim report, and table entries will represent the estimated total number of reporting aggregates of the given size that do some or all of their business in the given EPA region. There will not be any calculation of total dollar values by region and/or industry due to the categorical, open ended nature of questions 18 and 19.

Employee data (questions 10a-g and 11) will be profiled for each EPA region by industry category (see Tables 8 and 9). Table entries based on question 10 will represent the estimated total number of technical employees (questions 10a + 10b + 10d), the estimated total number of non-technical employees (questions 10c + 10e + 10f), the weighted number of reporting aggregates that the totals are based on, and the weighted number of reporting aggregates that did not report data. Table entries based on question 11 will represent estimated total number of certified or licensed employees, the weighted number of reporting aggregates that the total is based on, and the weighted number of reporting aggregates that did not provide answers to question 11.

Responses to questions about certification and licensing in multiple states (question 12 and 13) will be tabulated by industry category as shown in Tables 10 and 11. Further analyses of the data related to certification may be done in the future, but will not be included in the final report, due to the necessity of gathering detailed information on state regulatory programs.

MEMORANDUM

Page 4

February 21, 1984

Information on the number of locations associated with a reporting aggregate will be profiled by industry category (see Table 12). This analysis will be done in lieu of an analysis of the original responses to question 5 due to the considerable latitude of interpretations imposed on this question by respondents.

Each component of question 20 will be tabulated separately. Profiles of the percent data will be done for each industry category by the reporting aggregate size categories defined by question 18 (using the "collapsed" intervals), and will present the estimated total number of reporting aggregates in distinct "percent intervals" (see Table 13a). Profiles of the data on number of accounts will also be done for each industry category by the size categories, with each table cell containing the estimated total number of accounts, the weighted number of reporting aggregates that the total is based on, and the weighted number of reporting aggregates for which no account information was provided (see Table 14a).

Each section of question 21 will also be tabulated separately. Tables will profile the estimated total number of reporting aggregates in each of the percentage intervals for each of the reporting aggregate size categories (see Table 15a).

Service area of the reporting aggregates will be tabulated for each industry as the estimated total number of reporting aggregates serving categorized numbers of states or counties within states (see Table 16).

National trade association membership will be tabulated by size of reporting aggregate (see Table 17), as will information on subscriptions to trade journals (see Table 19). Regional trade association membership (see Table 18) and length of time in business (see Table 20) will be profiled by industry and size of reporting aggregate.

In order to assist the reader of the final report in assessing data quality, we may also provide weighted tabulations of responses to questions 23, 25, and 27. We lack to the ability to quickly quantify the various responses to question 26 because only the presence/absence of a response was coded on the computer file. However, some description of the type of complaints commonly reported (e.g., the questionnaire took too long to fill out, too hard to find or look up reg. numbers, bad timing, structure of the questionnaire does not fit the firm) will be included in the final report.

D. User File Content

Although agreement has been reached on the general type of data to be included on the user file, the actual level of detail and format of the data have yet to be decided upon. Maintenance of respondent confidentiality, the "generic" nature of some of the reg. numbers assigned to NUPAS data, ease of use, and minimization of possible user misinterpretation are factors that must be taken into account in such decisions. Based on our perceptions of EPA's needs and on the considerations listed above, we suggest that the user file have the following characteristics:

MEMORANDUM

Page 5

February 21, 1984

- each record on the file will represent one active ingredient in one product in one state reported on one questionnaire.
- the only direct linkage between records will be provided by a variable that indicates which other a.i. records (if any) are from the product record.
- a record will contain the following variables:
 1. "random" identification number,
 2. linkage identifier,
 3. a.i. name (trade name if possible),
 4. a.i. chemical code,
 5. class (i.e. herbicide, insecticide, etc.) of the product that contained the a.i.,
 6. formulation (collapsed version) of the product that contained the a.i.,
 7. whether or not the product was for restricted use,
 8. the state the a.i. was used in,
 9. the weighted amount of a.i. used in the state, in pounds,
 10. RTI's industry category variable,
 11. 7 variables representing the percents reported in question 20a-f,
 12. 5 variables representing the percents reported in question 21 (except for "other"), and
 13. national trade association membership of the respondent who reported the product.
 14. The percent by weight of active ingredient in the product, to one decimal place.

bkp

Table 1. Quantity of Active Ingredient Used in 1981, by EPA Region

Active Ingredient	EPA Chemical Code	EPA Regions										Total U.S.
		1	2	3	4	5	6	7	8	9	10	
<u>Insecticides</u>												
Chlordane	58201											
Methyl bromide	53201											
.												
.												
All others	-											
Total Insecticides												
<u>Herbicides</u>												
Betasan	9801											
.												
.												
.												
Total Herbicides												
<u>Fungicides</u>												
.												
.												
.												
Total Fungicides												
<u>Rodenticides</u>												
Warfin												
Talon												
.												
.												
.												
Total Rodenticides												
<u>Microbicides</u>												
.												
.												
.												
Total Microbicides												
<u>Overall</u>												
Chlordane												
.												
.												
.												
All others												

Note: Only the 10 most used a.i.'s (excluding general, nonspecific solvents) in each class would be listed. The class total would be for all a.i.'s in the class. Overall totals would be for the 10 most used active ingredients.

Table 2. Quantity of Active Ingredient Used in 1981, by Industry

Active Ingredient	Chemical Code	Lawn	Tree	Lawn and/or Tree	Structural	Structural and Lawn	Structural and Tree	Structural, Tree and/or Lawn
<u>Insecticides</u>								
Chlordane	58201							
Methyl bromide	53201							
.								
.								
All others								
Total Insecticides								
<u>Herbicides</u>								
Betasan	9801							
.								
.								
Total Herbicides								
<u>Fungicides</u>								
.								
.								
Total Fungicides								
<u>Rodenticides</u>								
Warfin								
Talon								
.								
.								
Total Rodenticides								
<u>Microbicides</u>								
.								
.								
Total Microbicides								
<u>Overall</u>								
Chlordane								
.								
.								
All others								

Note: Only the 10 most used a.i.'s (excluding general, nonspecific solvents) in each class would be listed. The class total would be for all a.i.'s in the class.

Table 3. Quantity of Active Ingredient Used in 1981, by Formulation

Chemical Type	Formulation	Active Ingredient	Chemical Code	Total Use (in lbs.)
Insecticide	Dust	Chlordane	58201	
		Sevin		
		.		
		.		
		<u>All others</u>		
		<u>Subtotal</u>		
	Wettable Powder	Chlordane		
		Sevin		
		.		
		.		
Herbicide		<u>All others</u>		
		<u>Subtotal</u>		
	Concentrate	.		
		.		
		<u>All others</u>		
		<u>Subtotal</u>		
	Wettable Powder	Betasan		
		Dacthal		
		.		
		.		
etc. Overall		<u>All others</u>		
		<u>Subtotal</u>		
	Concentrate	.		
		.		
		<u>All others</u>		
		<u>Subtotal</u>		
	Wettable powder	.		
		.		
		<u>All others</u>		
		<u>Subtotal</u>		

TABLE 4

FORMULATION CODES

CODE	DEFINITION
ø1	TECHNICAL CHEMICAL - Raw chemical ingredients as manufactured with no additives; usually 90% or greater active ingredient; ordinarily for formulating use only, but sometimes used directly as a ULV spray.
ø2	FORMULATION INTERMEDIATE - Technical chemical to which something has been added, e.g., a stabilizer; for formulating use only.
ø3	DUST - Active ingredient mixed with a powdered dry inert substance such as talc or clay; applied dry. Abbreviation: D
ø4	GRANDULAR - Vermiculite, attaclay, ground walnut shells, or other similar coarse particles impregnated with an active ingredient. Abbreviation: G
ø5	PELLETED/TABLETTED - Active ingredient mixed with binders, fillers and/or other inerts and formed into a pellet, tablet, or cake; also includes capsules (encapsulated material) which contain active ingredient alone or with inerts.
ø6	WETTABLE POWDER - Use as a suspension mixed in water. Abbreviation: WP or W
ø7	WETTABLE POWDER/DUST - Can be used either as a wettable powder in liquid suspension or dry as a dust.
ø8	CRYSTALLINE - An essentially pure chemical in solid form, such as copper sulfate (for water treatments) and paradichlorobenzene (moth crystals).
ø9	MICROENCAPSULATED - Time-release formulation.
1ø	IMPREGNATED MATERIALS - May be either a useful article or material impregnated with a pesticide (e.g., no-pest-strip, towellettes, weed-killing bars, roach tape).
12	EMULSIFIABLE CONCENTRATE - Always diluted for use; diluent is a liquid with polar characteristics opposite that of solvent in formulation; usually contains emulsifiers such as glycols, sulfonates, etc. Abbreviation: EC or E
13	INVERT EMULSION - Emulsion consisting of water droplets surrounded by oil (a common emulsion consists of oil droplets surrounded by water); this code should be used only when formulation is stated as an invert emulsion or invert emulsifiable; invert emulsions will not be encountered frequently.
14	FLOWABLE CONCENTRATE - A suspension of solid or semi-solid active ingredient in a liquid; always diluted for use; also called flowable solid.
15	SOLUBLE CONCENTRATE - Always diluted for use, forming a true solution; diluent is a solvent with the same polar characteristics as that of the materials in the formulation. Code products identified as 'spray concentrate' or 'liquid concentrate' as emulsifiable concentrate (12) unless information about the diluent and solvent is given such that you can decide specifically that it is a soluble concentrate (15).

(Continued)

<u>CODE</u>	<u>DEFINITION</u>
16	SOLUTION-READY TO USE - Used without dilution; may be a liquid, lotion, or paste. Abbreviation: S
17	OILS-NO ADDED PESTICIDE - Contains no active ingredient other than oils; generally petroleum distillates, mineral oil, hydrocarbon oils, creosote oil, etc.
18	PRESSURIZED GAS - Active ingredients are gaseous at room temperature and atmospheric pressure; packaged under pressure in a tank, or spray can; self-pressurized; sometimes called liquified gases.
19	PRESSURIZED LIQUID - Active ingredients are solid or liquid at room temperature and atmospheric pressure; packaged under pressure with appropriate solvents and propellants in a tank or spray can; released for use as an aerosol or liquid spray.
20	PRESSURIZED DUST - Active ingredient is a finely powdered solid or is mixed with a finely powdered inert such as talc; suspended in a volatile liquid under pressure in a tank or spray can; released for use as a dry powder spray.
90	FORMULATION NOT IDENTIFIED - This code should not be used except in special circumstances as approved by the coding supervisor.

Table 5a. Frequency Distribution of Original Formulation Codes in NUPAS Products¹

Formulation Code	Frequency
missing	23
1 ²	14
2 ²	6
3	56
4	172
5	46
6	101
7	12
8	4
9	2
10	14
12	245
14	29
15	166
16	93
17	12
18	7
19	68
20	1
90 ²	2
	<u>1073</u>

¹Because of the generic nature of reg. number assignment for some of the NUPAS data, the formulation codes are not and should not be interpreted as error-free representations of the formulations actually used by NUPAS respondents.

²Hard copy for products with these codes will be checked to see if other codes are more appropriate.

Table 5b. Frequency Distribution of Suggested Formulation
Categories for NUPAS Products

Formulation Category	Corresponding Formulation Codes ¹	Frequency
Dust	3,7	56 + 12 = 68
Granular/Pelleted	4,5	218
Wettable Powder	6,7	113 - 12 = 101
Liquid Concentrate	12,14,15	440
Ready to Use Solution	16	93
Pressurized	18,19,20	76
Other	missing, 1,2,8,9,10,17,90	77
		<u>1073</u>

¹See Table 4

Table 6. Gross Sales by Pest Control Firms in 1981 By EPA Region and Industry Category

EPA Region	Industry Category ¹	Number of Reporting Aggregates with Gross Sales of						
		Unknown	<\$25,000	\$ 25,000- \$100,000	\$100,000- \$ 500,000	\$ 500,000- \$1,000,000	\$1,000,000 \$5,000,000	>\$5,000,000
1	Tree/Lawn							
	Structural							
	Both							
	Overall							
2	Tree/Lawn							
	Structural							
	Both							
	Overall							
3	Tree/Lawn							
	Structural							
	Both							
	Overall							
etc. National	Tree/Lawn							
	Structural							
	Both							
	Overall							

¹It may be preferable to use the "expanded" industry categories.

Table 7. Sales of Pest Control Services in 1981 By EPA Region and Industry Category

EPA Region	Industry Category ¹	Number of Reporting Aggregates with Sales for Pest Control Services of				
		Unknown	<\$25,000	~~\$25,000- \$50,000~~	~~\$50,000- \$100,000~~	>\$100,000
1	Tree/Lawn					
	Structural					
	Both					
	Overall					
2	Tree/Lawn					
	Structural					
	Both					
	Overall					
3	Tree/Lawn					
	Structural					
	Both					
	Overall					
etc.						
National	Tree/Lawn					
	Structural					
	Both					
	Overall					

¹It may be preferable to use the "expanded" industry categories.

Table 8. Number of Employees of Pest Control Firms in 1981
By EPA Region and Industry Category

EPA Region	Industry Category ¹	Number of		Weighted Number of reporting aggregates that provided data	Weighted Number of reporting aggregates that did not Provide Data
		Service Technicians and Technical Managers	Clerical, Sales, and Other Employees		
1	Tree/Lawn Structural Both Overall				
2	Tree/Lawn Structural Both Overall				
	.				
	.				
	.				
etc.					
National	Tree/Lawn Structural Both Overall				

¹It may be preferable to use "expanded" industry categories.

Note: Respondent misinterpretation and the lack of assurance that responses were at the reporting aggregate level make data quality suspect for this table.

Table 9. Number of Certified and Licensed Employees of Pest Control Firms in 1981,
By EPA Region and Industry Category

EPA Region	Industry Category ¹	Number of Certified and Licensed Employees	Weighted Number of Reporting Aggregates That Provided Data	Weighted Number of Reporting Aggregates That Did Not Provide Date
1	Tree/Lawn Structural Both Overall			
2	Tree/Lawn Structural Both Overall			

etc.

¹It may be preferable to use the "expanded" industry categories.

Note: Respondent misinterpretation and the lack of assurance in nontelephone states that responses were at the reporting aggregate level make data quality suspect for this table.

Table 10. Multiple State Certification or Licensing, By Industry Category

Industry Category ¹	Have Employees Who Are Certified or Licensed in More Than One State	Do Not Have Employees Who Are Certified or Licensed in More Than One State	Don't Know if Any of Their Employees Are Certified or Licensed in More Than One State	Did not Provide Data
Tree/Lawn				
Structural				
Both				
Overall				

¹It may be preferable to use "expanded" categories.

Table 11. Methods of Multiple State Certification or Licensing¹, By Industry Category

Industry Category ²	Formal Examination	Reciprocity	Examination and Reciprocity	Other or Unknown	No Response
Tree/Lawn					
Structural					
Both					
Overall					

¹Only reporting aggregates that reported having any employees certified or licensed in multiple states are included in this analysis.

²It may be preferable to use "expanded" categories.

Table 12. Number of Business Locations Associated With Reporting Aggregate,
By Industry Category

Industry Category	Service Area	Number of Business Locations in Reporting Aggregate						
		1	2-3	4-5	6-10	11-25	26-50	>50
Tree	Single state							
	Multi-state							
Lawn	Single state							
	Multi-state							
Tree and/or Lawn	Single state							
	Multi-state							
Structural	Single state							
	Multi-state							
Structural & Tree	Single state							
	Multi-state							
Structural & Lawn	Single state							
	Multi-state							
Structural and Tree and/or Lawn	Single state							
	Multi-state							
Overall	Single state							
	Multi-state							

Table 13a. Percent of Gross Sales From Pest Control Services at Residential Single Family Structures By Industry Category and Reporting Aggregate Size

Industry Category	Reporting Aggregate Size gross sales, in dollars)	Percent of Sales Associated with Site					Data missing
		0	1-25	26-50	51-75	76-99	100
Tree/Lawn	<25,000						
	25,000-100,000						
	100,000-500,000						
	500,000-1,000,000						
	1,000,000-5,000,000						
	>5,000,000						
Structural	Data missing						
	<25,000						
	25,000-100,000						
	100,000-500,000						
	500,000-1,000,000						
	1,000,000-5,000,000						
Both	>5,000,000						
	Data missing						
	<25,000						
	25,000-100,000						
	100,000-500,000						
	500,000-1,000,000						
Overall	1,000,000-5,000,000						
	>5,000,000						
	Data missing						
	<25,000						
	25,000-100,000						
	100,000-500,000						

Note: Separate tables would be prepared for questions 20a², b,c,d,e, and f.

Note: Respondent error and/or misinterpretation make data quality for all question 20 tables suspect.

Table 14a. Number of Accounts at Residential Single Family Structures
By Industry Category and Reporting Aggregate Size

Industry Category	Reporting Aggregate Size (gross sales, in dollars)	Number of Accounts	Number of Reporting Aggregates That Reported Data	Number of Reporting Aggregates For Which Data Was Not Provided
Tree/Lawn	<25,000			
	25,000-100,000			
	100,000-500,000			
	500,000-1,000,000			
	1,000,000-5,000,000			
	>5,000,000			
	Data missing			
Structural	<25,000			
	25,000-100,000			
	100,000-500,000			
	500,000-1,000,000			
	1,000,000-5,000,000			
	>5,000,000			
	Data missing			
Both	<25,000			
	25,000-100,000			
	100,000-500,000			
	500,000-1,000,000			
	1,000,000-5,000,000			
	>5,000,000			
	Data missing			
Overall	<25,000			
	25,000-100,000			
	100,000-500,000			
	500,000-1,000,000			
	1,000,000-5,000,000			
	>5,000,000			
	Data missing			

Note: Separate tables would be prepared for questions 20a², b,c,d,e, and f.

Note: Respondent error and/or misinterpretation make data quality for all question 20 tables suspect.

Table 15a. Percent of Gross Sales from Lawn and Turf By Reporting Aggregate Size

Reporting Aggregate Size gross sales, in dollars)	Percent of Gross Sales Associated with Lawn and Turf						Data missing
	0	1-25	26-50	51-75	76-99	100	
<25,000							
25,000-100,000							
100,000-500,000							
500,000-1,000,000							
1,000,000-5,000,000							
>5,000,000							
Data missing							

Note: Separate tables would be prepared for percent of gross sales from tree, shrub, and ornamental, termites, commodity fumigation, and general pest control.

Table 16. Service Area of Reporting Aggregate, By Industry Category

Service Area of the Reporting Aggregate	Industry Category			Overall
	Tree/Lawn	Structural	Both	
Multiple States				
2 - 3				
4 - 5				
5 - 10				
11 - 25				
26 - 50				
Single state, ~~all~~ counties				
Single state, multiple counties				
Single state, single county				
Don't know				
Data missing				

Table 17. National Trade Association Membership, by Size of Reporting Aggregate

Reporting Aggregate Size (Gross Sales, in dollars)	Membership in National Associations				
	None	NPCA	NAA	PLCAA	No Response
<25,000					
25,000-100,000					
100,000-500,000					
500,000-1,000,000					
1,000,000-5,000,000					
>5,000,000					
Data missing					

Table 18. Membership in State or Regional Trade Associations,
by Industry Category and Size of Reporting Aggregate

Industry Category	Reporting Aggregate Size (gross sales, in dollars)	Membership in State or Regional Trade Associations			
		Yes	No	Don't know	Data missing
Tree/Lawn	<25,000				
	25,000 - 100,000				
	.				
etc.	.				
	.				
	.				
Overall	<25,000				
	25,000 - 100,000				
	.				
	.				
	.				

Table 19. Subscriptions to Trade Journals, By Reporting Aggregate Size

Subscriptions	<25,000	25,000- 100,000	100,000- 500,000	500,000- 1,000,000	1,000,000- 5,000,000	>5,000,000	Data Missing
None							
American Lawn Applicator							
Arborage							
Grounds Maintenance							
Landscape & Turf							
Landscape Industry							
Lawn Care Industry							
Pest Control							
Pest Control Technology							
Pest Management							
Weeds, Trees, and Turf							
Non Response							

Table 20. Number of Years in Business, by Industry Category and Size of Reporting Aggregate

Category	Reporting Aggregate Size (gross sales, in dollars)	Number of years in Business					No response
		<1	1-2	3-5	5-10	10-25	
Tree/Lawn	<25,000						
	25,000 - 100,000						
	100,000 - 500,000						
	500,000 - 1,000,000						
	1,000,000 - 5,000,000						
	>5,000,000						
	Data Missing						
Structural	<25,000						
	25,000 - 100,000						
	100,000 - 500,000						
	500,000 - 1,000,000						
	1,000,000 - 5,000,000						
	>5,000,000						
	Data Missing						
Both	<25,000						
	25,000 - 100,000						
	100,000 - 500,000						
	500,000 - 1,000,000						
	1,000,000 - 5,000,000						
	>5,000,000						
	Data Missing						
Overall	<25,000						
	25,000 - 100,000						
	100,000 - 500,000						
	500,000 - 1,000,000						
	1,000,000 - 5,000,000						
	>5,000,000						
	Data Missing						